STRUCTURAL MODELS OF BACTERIAL RNA CHAPERONE HFQ IN COMPLEX WITH THE RNA TARGETS

By
Penny Yi Peng

A dissertation submitted to Johns Hopkins University in conformity
with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
November 2014

Abstract

Bacterial Sm protein Hfq is required for translation regulation by small non-coding RNAs (sRNAs), which play a key role in bacterial environmental adaptation and pathogenic pathways. Hfq promotes anti-sense base pairing of the sRNAs that inhibits or activates translation of the mRNA target. For the positive regulation of stress response regulator *rpoS*, Hfq distal surface binds to an upstream (AAN) motif in the mRNA leader, while the proximal surface and lateral rim bind to the U-rich sequences in the sRNAs.

I found that the Hfq activity requires the (AAN) motif located < 80 nt upstream of the sRNA annealing site in the *rpoS* leader, because Hfq lateral rim directly binds a downstream U_5 motif, as shown by SHAPE, in vivo *LacZ* activity assays, and in vitro binding assays. Small angle X ray scattering (SAXS) showed that Hfq folds the *rpoS* mRNA leader into a compact tertiary conformation. Finally, I generated all-atom structural models of *rpoS*•Hfq complex by combining MC-Sym structure prediction, rigid body modeling, and Monte Carlo simulation based on the SAXS data. The dynamic interaction models suggest that Hfq binds the U_5 motif to position and unwind the inhibitory stem for sRNA entry and annealing, and that Hfq disengages from the downstream domain to release the RNA duplex for translation initiation.

I also discovered that Hfq does not bind a U-rich motif in the *flhD* leader or restructures the RNA significantly, causing only moderate enhancement for sRNA annealing efficiency. Moreover, Hfq binds to various sRNAs by distinctive modes depending on the presence of the A-rich or U-rich sequences; however, annealing activity relies more on the complementarity between the base pairing RNAs than the binding mode for a specific RNA.

Advisory Committee

Professor Sarah A. Woodson, Advisor and Primary Reader

Professor Karen Beemon, Secondary Reader

Professor Joseph G. Gall

Professor Xin Chen

Dedication

To my parents and husband

Acknowledgement

I would like to thank my advisor Dr. Sarah Woodson for training me to be a scientist; I will always cherish those aha moments when we solved a scientific puzzle together. I would like to thank Dr. Toby Soper and Dr. Subrata Panja for guiding me into the Hfq research; Dr. Hui-Ting Lee, Dr. Sanjaya Abeysirigunawardena and Dr. Duncan Kilburn for sharing their insightful ideas; and all the other Woodson lab members for creating this intellectually nourishing environment.

I am grateful for meeting so many amazing scientists in the community. I would like to thank Dr. Susan Gottesman, Dr. Gisela Storz, and all the members in their laboratories for the valuable discussions. I especially thank Dr. Nadim Majdalani for the generous offer of reagents and protocols. I would like to thank our collaborators Dr. Xianyang Fang and Dr. Joseph Curtis for lending me their expertise to move the project forward.

I particularly thank my thesis committee members Dr. Joseph Gall and Dr. Karen Beemon. Thank you for caring me along the way and thank you for supporting my professional and personal development.

I would like to give my deepest thanks to my family. I would like to thank my parents for their unconditional love and constant encouragement. I would like to thank my husband for giving me everything I could ever ask for along this journey. I would like to thank my family-in-law for providing me a home in Canada.

Finally, I would like to thank everyone who helps me, past and present, for making me a stronger person to stand up against any challenges in my life.

Table of Contents

List of Figures and Tables

Chapter 1 Introduction

Hfq belongs to the Sm protein superfamily

Sm superfamily members share similar structures

Human Sm proteins were originally discovered in the 1960s as the antigens in the Sm (Stephanie Smith) serotype of patients who suffered the autoimmune disease, systemic lupus erythematosus (SLE) (Tan and Kunkel, 1966). Extensive research in the following decades revealed the ubiquitous presence of the Sm superfamily in all three domains of life, as well as the fascinating details of Sm protein structure, assembly and function (Mura et al., 2013).

All Sm proteins share a common fold, in which the peptide backbone deviates very little as reflected by the root-mean-squared deviation (RMSD) of the atomic coordinates. This structural similarity suggests that a highly conserved "Sm fold" was present in the universal common ancestor before the divergence of bacteria, archaea, and eukaryotes (Mura et al., 2013). The Sm fold (β1-5), capped by an N-terminal α-helix and separated by a variable linker, creates a subunit interface via the outer β-sheets (β4 and β5) that facilitate oligomerization (Fig. 1.1) (Weichenrieder, 2014).

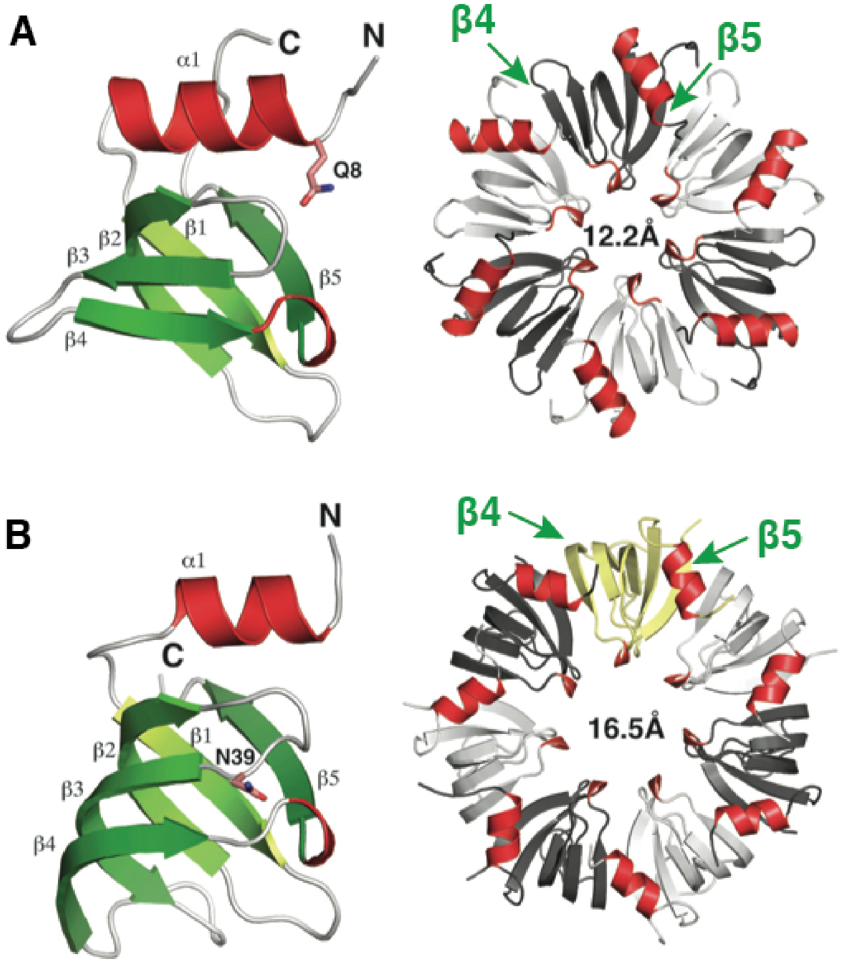


Figure 1.1. Crystal Structures of ring-shaped (L)Sm and Hfq proteins (adapted from "RNA binding by Hfq and ring-forming (L)Sm proteins: a trade-off between optimal sequence readout and RNA backbone conformation") (Weichenrieder, 2014). (*A*) Bacterial *St*Hfq and (*B*) Archaeal AfSm1 proteins self-assemble into ring-shaped hexamer and heptamer, respectively. Left, each monomer features conserved Sm1 and Sm2 motifs (β1-3 and β4-5, colored green) separated by a variable linker (loop connecting β3-4). The entire Sm fold is capped by an extended N-terminal α-helix (red). Right, protomers are colored white, grey, and yellow, and the N-terminal α-helix is red; β4 and β5 are indicated to show the subunit interface.

This self-assembly capability leads to a ring-shaped quaternary structure observed for all Sm proteins (Mura et al., 2013; Weichenrieder, 2014). For instance, bacterial Sm proteins (Hfq) mostly form homohexameric rings (Fig. 1.1A) (Horstmann et al., 2012; Kadowaki et al., 2012; Kovach et al., 2014; Link et al., 2009; Schumacher et al., 2002; Someya et al., 2012). By contrast, archaeal and eukaryotic Sm proteins assemble into homo- and hetero-heptameric rings, respectively (Fig. 1.1B) (Collins et al., 2001; Leung et al., 2011; Mura et al., 2001; Pomeranz Krummel et al., 2009; Törö et al., 2001).

Sm proteins bind RNAs at the Sm folds

The subunit interface of the Sm folds acts as a conserved binding pocket for U-rich RNAs, which is present across the entire Sm superfamily, although the RNA-protein interacting geometry is slightly different in different family members. In addition, bacterial Hfq recognizes A-rich as well as U-rich RNAs on its distal and proximal surfaces (Fig. 1.4 and 1.5) (Horstmann et al., 2012; Kadowaki et al., 2012; Kovach et al., 2014; Link et al., 2009; Schumacher et al., 2002; Someya et al., 2012). By contrast, archaeal Sm proteins only bind U-rich RNAs on the proximal surface (Collins et al., 2001; Mura et al., 2001; Törö et al., 2001). Eukaryotic Sm proteins bind U-rich RNAs in a more sophisticated and complex conformation. In the example of U4 snRNAP, the heptameric Sm ring contains a funnel shaped central hole lined with RNA-binding motifs, and the snRNA threads through this central hole (Fig. 1.2) (Leung et al., 2011).

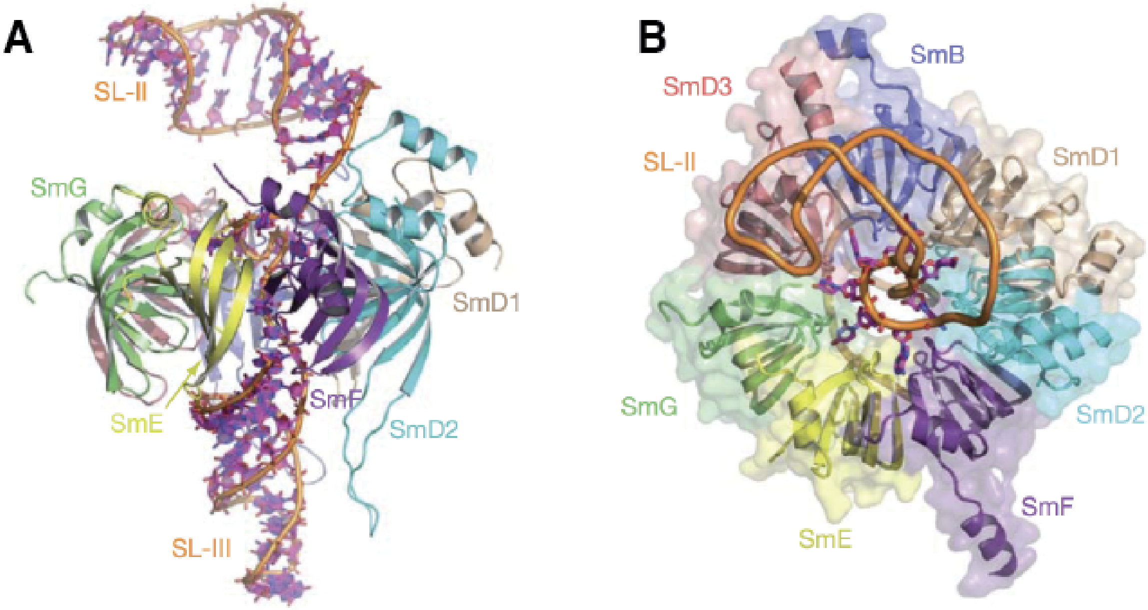


Figure 1.2. Structure of the U4 snRNP core domain showing the snRNA threading through the central hole of the Sm ring (reprinted from "Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis") (Leung et al., 2011). (*A*) Side view of the core domain: the heptameric Sm ring shows a funnel shape with the flat proximal face up and the tapered distal face down; the Sm folds line the walls of the central hole to interact with the SnRNA. (*B*) Top view of the flat face showing nucleotides bound near the inner rim of the proximal surface.

From an evolutionary perspective, the increasing complexity of Sm assembly and RNA-binding not only accommodates the increasing requirement for Sm-based RNA processing; it also reflects a gradual shift of Sm functions from a versatile chaperone to a specific scaffold (Fig. 1.3).

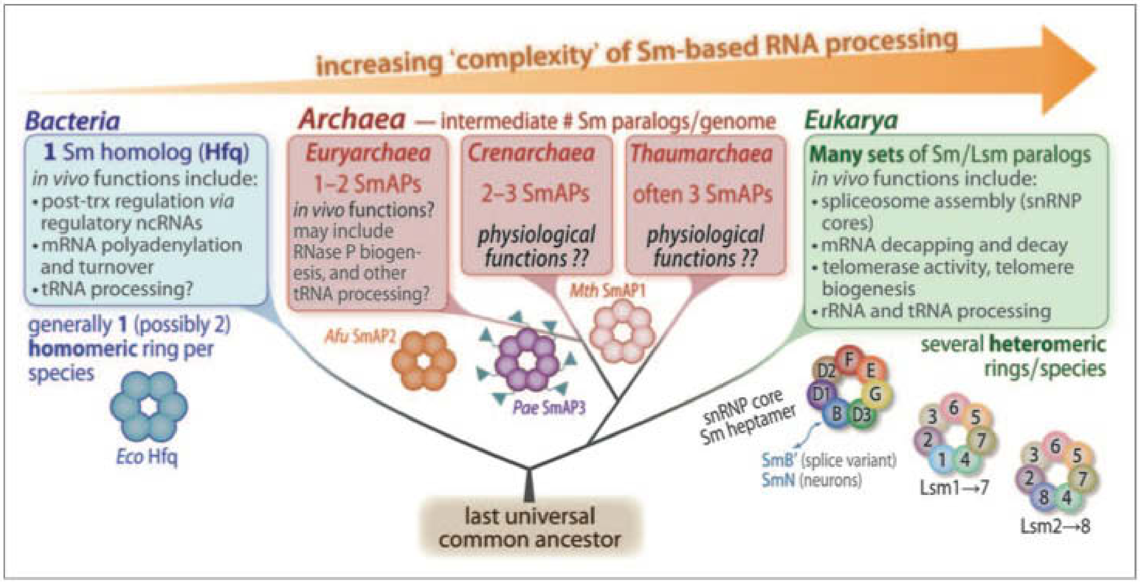


Figure 1.3. Evolution of Sm protein assembly and function (reprinted from "Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function") (Mura et al., 2013). This phylogenetic tree illustrates the presence of the Sm protein superfamily in Bacteria, Archaea, and Eukarya. Sm assembly gradually evolved from homohexamers in bacterial Sm proteins (Hfq) to heteroheptamers in eukaryotes. While this correlates with the increasing complexity of Sm-based RNA processing, it also illustrated a shift of Sm function from RNA chaperones to snRNP scaffolds.

Sm proteins regulate RNA-related cellular functions

As an RNA chaperone, the bacterial branch of the Sm superfamily (Hfq) regulates a broad spectrum of physiological activities (Mura et al., 2013). Extensive research has explored the phenotypes of *hfq* mutations in various bacterial species, from human pathogens to plant symbionts (Sobrero and Valverde, 2012). In most cases, the absence of Hfq compromises the fitness of the bacterial cells in responses to environmental stresses (Gottesman et al., 2006) such as iron limitation (Masse and Gottesman, 2002) and osmotic shock (Guillier and Gottesman, 2006). Moreover, *hfq* null strains of bacterial pathogens are more sensitive to host defense mechanisms and produce highly attenuated infections in various animal models, suggesting that Hfq may control virulence factor expression and pathogenic secretion systems (Chao and Vogel, 2010).

By contrast, eukaryotic Sm/LSm proteins function as a molecular scaffold for spliceosome assembly as well as regulating RNA metabolism. Sm proteins (SmB/B’, D3, D2, D1, E, F, and G) and LSm proteins (LSm2-8) recruit specific snRNAs into highly organized snRNPs that process pre-mRNAs into mature forms (Will and Luhrmann, 2001). Another LSm heptameric complex (LSm1-7) has been reported to facilitate mRNA decapping during 5’-to-3’ mRNA decay (Tharun et al., 2000).

Archaeal Sm proteins fall into two subfamilies (Sm1 and Sm2) with poorly understood functions. They may have activities similar to Hfq for regulating translation or RNA metabolism by interacting with small RNAs and other protein factors (Murina and Nikulin, 2011). Despite their closer structural homology to eukaryotic Sm proteins, archaeal Sm are unlikely to be involved in splicing due to the lack of an archaeal spliceosome.

Taken together, Sm proteins engage in a remarkable range of RNA-related functions across bacterial, archaeal, and eukaryotic kingdoms. Among all the Sm proteins, Hfq achieve the greatest breadth of activities by interacting with a variety of nucleic acids and proteins. Understanding how Hfq interacts with its constellation of targets not only enlightens the evolutionary perspectives of the Sm superfamily, but also elucidates the underlying mechanisms of the bacterial stress response and pathogenic pathways.

Hfq binds A-rich and U-rich RNAs

Discovery of Hfq

Hfq was first identified as host factor I (HF-I) required for the replication of bacteriophage Qβ RNA in the 1960s (Franze de Fernandez et al., 1968), and shown to be a hexameric protein able to bind single-stranded RNAs (Franze de Fernandez et al., 1972). Subsequent studies in the 1980s focused on HF-I's nucleic acid binding properties and revealed its high affinity for A-rich sequences (de Haseth and Uhlenbeck, 1980; de Haseth and Uhlenbeck, 1980; Senear and Steitz, 1976). After the *E. coli* gene *hfq* encoding HF-I was cloned and sequenced in 1991 (Kajitani and Ishihama, 1991), *hfq* deletions uncovered pronounced pleiotropic phenotypes associated with the absence of HF-I (Tsui et al., 1994).

However, Hfq's roles in gene regulation remained a puzzle until 1996, when independent studies confirmed the requirement of Hfq for efficient expression of the global regulatory factor RpoS in *E.coli* (Brown and Elliott, 1996; Muffler et al., 1996). This led to a new chapter of Hfq research, where regulatory small RNAs (sRNAs) including DsrA, RprA, Spot42, OxyS and RyhB were also found to interact with Hfq (Masse and Gottesman, 2002; Moller et al., 2002; Sledjeski et al., 2001; Zhang et al., 1998). With the advent of high-throughput sequencing techniques, dozens of sRNAs were identified in a variety of bacteria species (Christiansen et al., 2006; Dambach et al., 2013; Sittka et al., 2008; Zhang et al., 2003). The most recent Hfq-binding transcriptome study detected almost 300 mRNAs, 25 intragenic RNAs (intraRNAs) and 67 antisense RNAs (asRNAs) in *E.coli* (Bilusic et al., 2014).

Crystal structures of Hfq core

In parallel, the structural basis for Hfq's RNA-binding property has been gradually established since 2002, through the use of comparative structural modeling (Arluison et al., 2002; Sun et al., 2002), electron microscopy (Zhang et al., 2002), and crystallography (Schumacher et al., 2002). As of this writing, 40 crystal structures have been deposited on the PDB server: 11 from *Escherichia coli* (*Ec*Hfq) (Beich-Frandsen et al., 2011; Hammerle et al., 2012; Link et al., 2009; Robinson et al., 2014; Sauter et al., 2003; Wang et al., 2011; Wang et al., 2013; Yonekura et al., 2013), 2 from *Staphylococcus aureus* (*Sa*Hfq) (Schumacher et al., 2002), 2 from *Salmonella typhimurium* (*St*Hfq) (Sauer and Weichenrieder, 2011), 11 from *Pseudomonas* *aeruginosa* (*Pa*Hfq) (Moskaleva et al., 2010; Murina et al., 2013; Murina et al., 2014; Nikulin et al., 2005), and 14 from other bacteria species (Baba et al., 2010; Boggild et al., 2009; Kadowaki et al., 2012; Kovach et al., 2014; Nielsen et al., 2007). Deciphering the similarities and differences of various Hfq structures provides invaluable insights for Hfq-RNA interaction mechanisms.

The Hfq subunit interface at the center of the proximal surface serves as an RNA binding pocket that preferably targets U-rich RNAs, but can also tolerate C- and A-nucleotides (Fig. 1.4) (Weichenrieder, 2014). This pocket prefers the 3'-ends of RNAs for optimal backbone geometry (Weichenrieder, 2014). For instance, the *Sa*Hfq-AU$_5$G complex showed a deflated oligonucleotide conformation and an expelled 3'-terminal G-nucleotide (Fig. 1.4A) (Schumacher et al., 2002). By contrast, *St*Hfq bound U$_6$ in a constricted conformation with the 3'-hydroxyl group directly recognized by H57 residue at the binding pocket (Fig. 1.4B) (Sauer and Weichenrieder, 2011). Furthermore, *Ec*Hfq not only accommodated a 3'-terminal A-nucleotide in a manner similar to *St*Hfq, but also engaged Mg^{2+} to stabilize the irregular RNA conformation (Fig. 1.4C) (Wang et al., 2011).

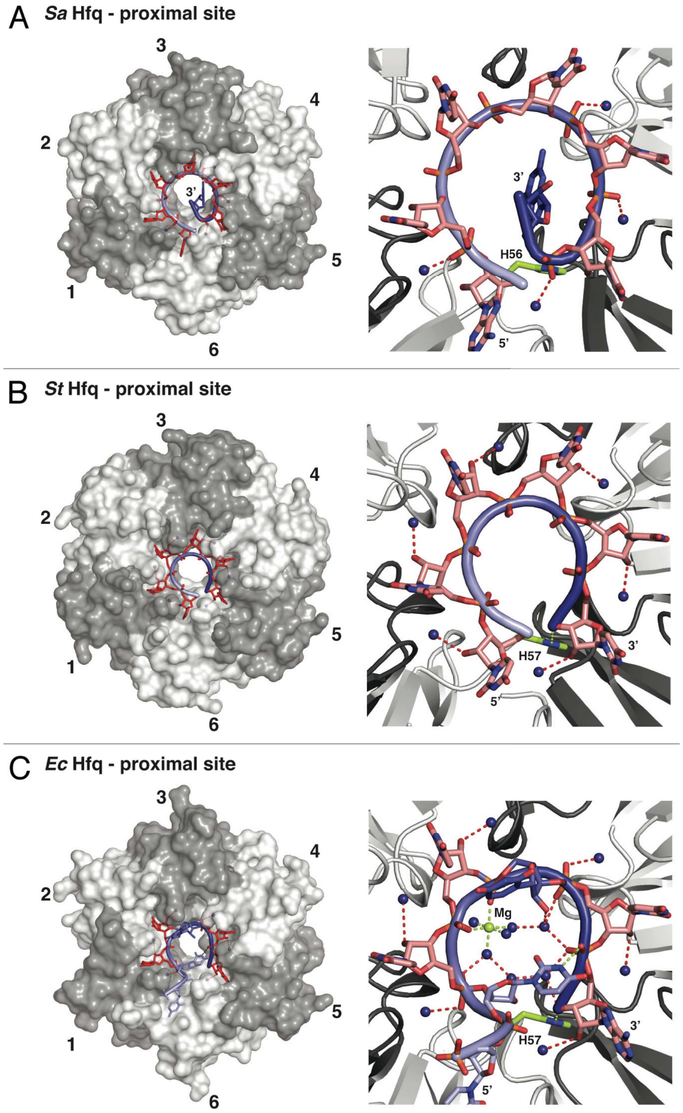


Figure 1.4. Crystal structures of bacterial Hfq showing U-rich RNAs binding to the proximal surface (reprinted from "RNA binding by Hfq and ring-forming (L)Sm proteins: a trade-off between optimal sequence readout and RNA backbone conformation") (Weichenrieder, 2014). (*A*) *Sa*Hfq in complex with rAU_5G oligonucleotide: the RNA conformation is dilated and the 3'-nucleotide is expelled from the ring (colored in blue). (*B*) *St*Hfq in complex with rU_6 oligonucleotide: the constricted RNA conformation allows the recognition of the 3'-end by H57 (colored in lime). (*C*) *Ec*Hfq in complex with rAU_6A oligonucleotide: the irregular RNA backbone alternates between the dilated and constricted conformations that are stabilized by Mg^{2+} ions (colored in lime).

Intriguingly, the distal face of Hfq provides another RNA binding site (Mikulecky et al., 2004), which specifically recognizes A-nucleotides (Fig. 1.5). Unlike the highly conserved proximal binding pocket, the distal face binding varies more among bacteria species. In gram-negative bacteria, a $(AAN)_n$ motif sits on the Hfq distal surface like a wavy crown because the first two A-nucleotides fit into the “A” site and a shallow pocket, whereas the third nucleotide bulges outwards (Fig. 1.5A) (Link et al., 2009). In contrast, Hfq in gram-positive bacteria lack the additional pocket and recognize a $(AN)_n$ motif (Fig. 1.5B) (Horstmann et al., 2012; Someya et al., 2012).

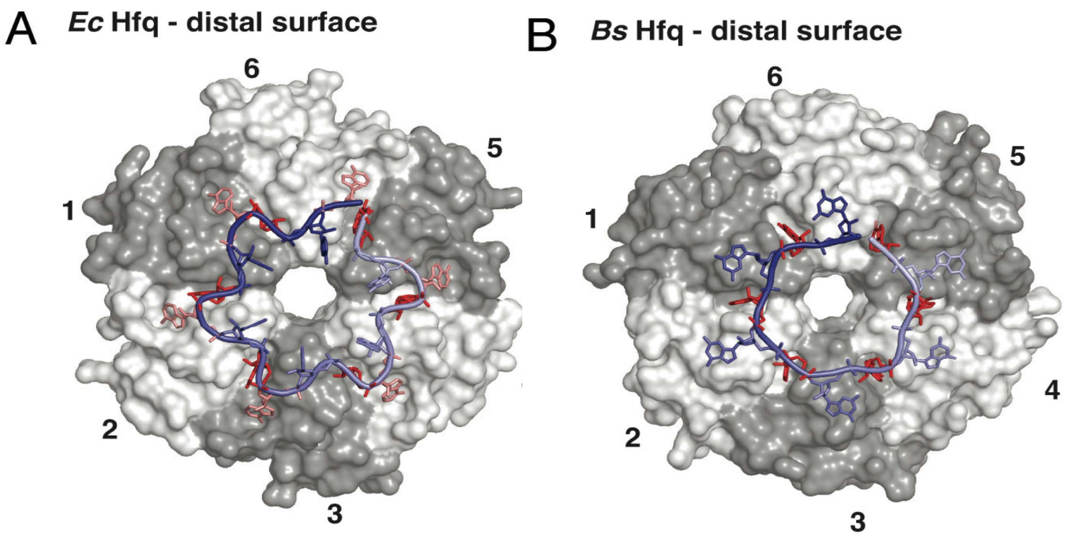


Figure 1.5. Crystal structures of bacterial Hfq showing A-rich RNAs binding to the distal surface (reprinted from “RNA binding by Hfq and ring-forming (L)Sm proteins: a trade-off between optimal sequence readout and RNA backbone conformation”) (Weichenrieder, 2014). (*A*) *Ec*Hfq in complex with r(AAN)$_n$ oligonucleotide: AAN triplet nucleotides are colored in dark red, light red, and blue. (*B*) *Bs*Hfq in complex with r(AN)$_n$ oligonucleotide: AN duplet nucleotides are colored in dark red and blue.

Additional structural features for RNA-interaction

The Hfq lateral rim was only recently identified as a third RNA binding site, which finally explained the enigma why Hfq binds sRNAs even when 3'-recognition is blocked (Fig. 1.6) (Weichenrieder, 2014). This site is an arginine-rich basic patch that is highly conserved in gram-negative bacteria (Sauer and Weichenrieder, 2011; Wang et al., 2011). Although direct crystal structure has not been reported, *Pa*Hfq and *Pa*Sm1 show additional U-binding sites that are close to this basic patch when superimposed with *Ec*Hfq (Murina et al., 2013; Thore et al., 2003). Nevertheless, the lateral surface is significant for Hfq function, because the rim directly associates with the sRNA bodies to stabilize the transient stem-loop during structural remodeling or strand annealing (Ishikawa et al., 2012; Sauer et al., 2012).

Finally, Hfq C-termini may also assist Hfq regulation, although the evidence remains inconclusive (Fig. 1.6). Unlike the conserved core, the Hfq C-terminus displays considerable variation in length and sequences among bacterial species (Sun et al., 2002). Functional studies shows that *Ec*Hfq C-terminal truncation binds sRNAs but not mRNAs (Vecerek et al., 2008), coincident with the evidence that most mRNA-sRNA regulation has been found in the species bearing Hfq proteins with long C-termini (Jousselin et al., 2009). However, the smallest known Hfq (found in the archeon *Methanocaldococcus jannaschii*) promotes *E.coli* sRNAs intermolecular base pairing *in vitro* and compensates some phenotypes displayed by endogenous *hfq* knockout strains (Nielsen et al., 2007), suggesting that the C-terminus may be dispensable for Hfq activity. From a structural perspective, the disordered nature of the Hfq C-terminus (Rajkowitsch and Schroeder, 2007; Vecerek et al., 2008) is a common feature among many RNA chaperones (Tompa and Csermely, 2004). In fact, the presence of C-terminus affected crystal packing of *Ec*Hfq, although the disordered portion was not solved in the final structure (Beich-Frandsen et al., 2011).

With the accumulating structural details of Hfq-RNA interaction, it becomes clear that Hfq targets a vast diversity of RNAs by recognizing the A-rich and U-rich sequences on its multiple surfaces. Unsurprisingly, this extraordinary versatility allows Hfq to regulate a broad array of cellular activities where RNA is involved.

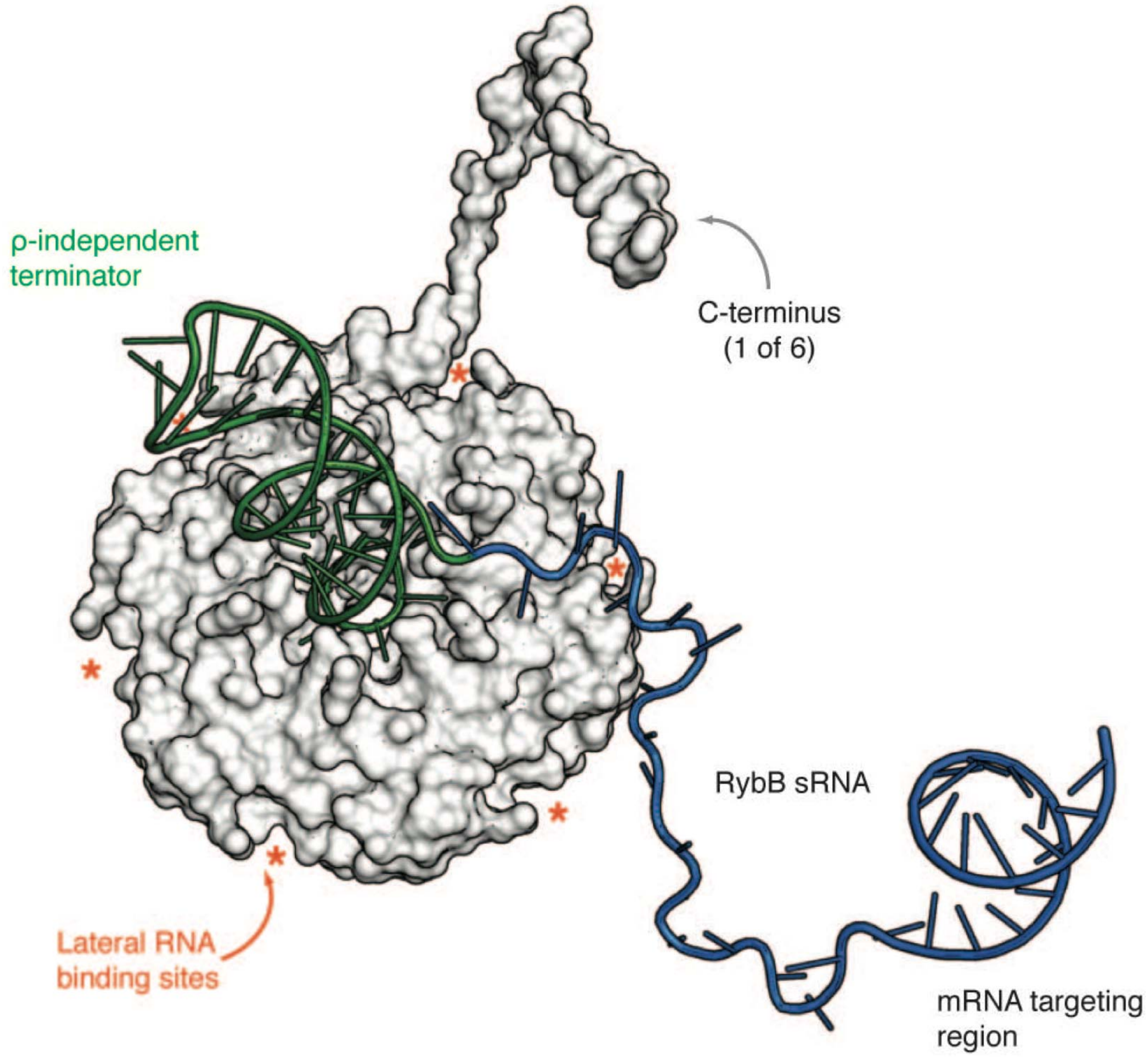


Figure 1.6. Model of an Hfq-sRNA complex showing additional RNA interacting surface of Hfq (reprinted from "Structure and RNA binding properties of the bacterial Sm protein Hfq") (Sauer, 2013). Hfq (PDB-ID: 2YLC) is shown as surface representation with a superimposed model of one C-terminus for clarity. While the 3'-U tail of the sRNA (green) binds Hfq proximal surface, the U-rich patches of the sRNA body (blue) directly interact with Hfq lateral surface (binding sites indicated by orange asterisks).

Hfq riboregulation in action

Bacterial sRNAs regulate mRNA translation

Enterobacterial cells encode approximately 200-300 small non-coding RNAs (sRNAs) that corresponds to ~5% of the total genomic genes (Storz et al., 2004). These sRNAs are generally transcribed as ~100 nt entities with a few stem-loop structures (Gottesman and Storz, 2011). They fine-tune gene expression for different growth environments, allowing the bacteria to survive various stress conditions (Hoe et al., 2013). Moreover, sRNAs were found to regulate microbial pathogenesis (Michaux et al., 2014).

Similar to eukaryotic microRNAs (miRNAs) and small interfering RNAs (siRNAs), bacterial sRNAs also base pair with the mRNA targets to regulate translation and RNA stability (Gottesman and Storz, 2011). For most sRNAs that are encoded at a different locus than the mRNA targets, complementarity is limited and generally contains at least a seed region of 6-8 contiguous base pairs (Gottesman and Storz, 2011).

In most cases, sRNAs anneal at the ribosome binding site of the mRNA leaders, and repress translation initiation by preventing ribosome entry (Fig. 1.7A). In other examples, sRNA binding activates translation initiation by resolving the mRNAs inhibitory secondary structure that otherwise cover the ribosome binding site (Fig. 1.7B).

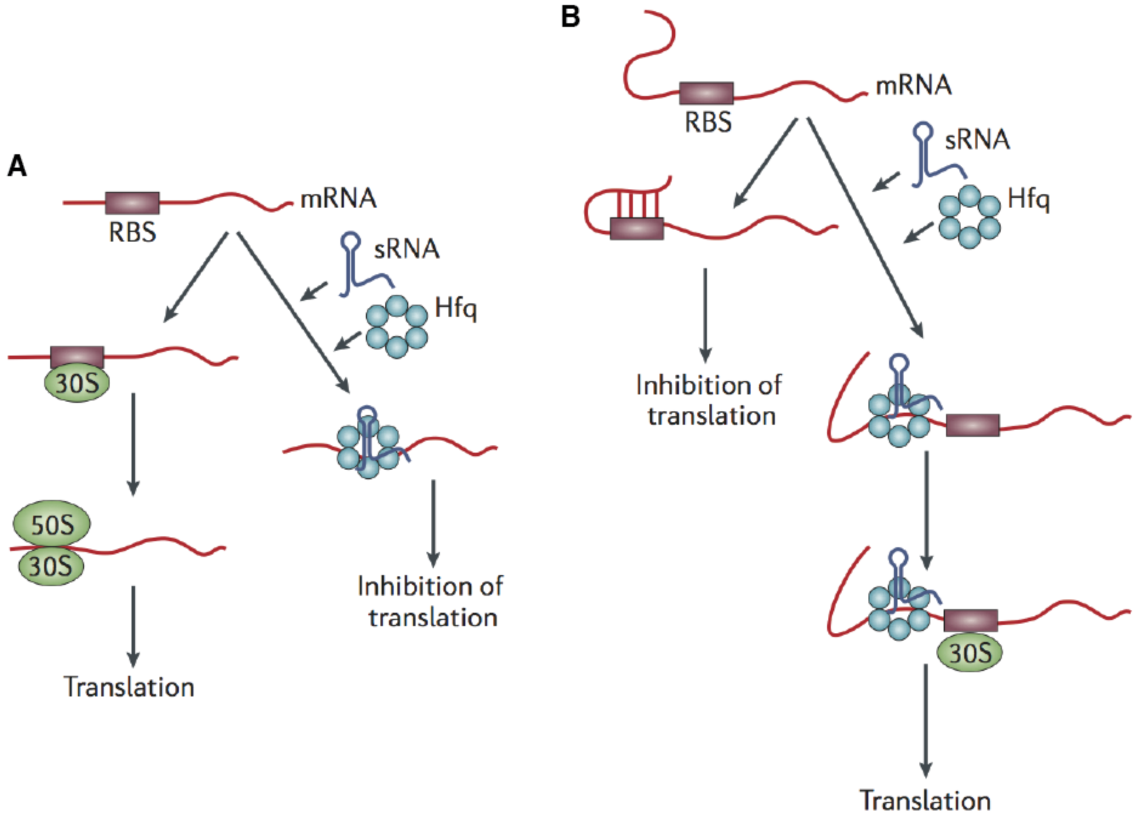


Figure 1.7. Widely accepted models for sRNA regulation of translation (adapted from "Hfq and its constellation of RNA") (Vogel and Luisi, 2011). (*A*) sRNA anneals to the ribosome binding site in the mRNA leader and represses translation initiation. (*B*) sRNA binding opens up an inhibitory secondary structure of the mRNA, allowing ribosome entry and activating translation initiation. In both cases, Hfq facilitates sRNA-mRNA annealing.

Hfq regulates *rpoS* translation

A well-studied example of positive regulation by sRNAs, *rpoS* encodes an alternative stationary phase sigma factor that mediates the expression of many stress response genes (Fig. 1.8) (Hengge-Aronis, 2002). During exponential growth, translation of the *rpoS* mRNA is inhibited by a stable stem-loop that masks the Shine-Dalgarno sequence (Fig. 1.8B) (Majdalani et al., 1998). Three *E. coli* sRNAs (DsrA, RprA, and ArcZ) up regulate *rpoS* translation by base pairing with the *rpoS* leader and opening the inhibitory stem, releasing the Shine-Dalgarno sequence for translation initiation (Battesti et al., 2011).

Hfq mediates effective annealing of activating sRNAs to *rpoS* mRNA through an $(AAN)_4$ repeat far upstream within the rpoS 5’ UTR (Fig. 1.9) (Soper et al., 2010; Soper and Woodson, 2008). The resulting ternary complex not only opens the inhibitory secondary structure, but also stabilizes *rpoS* mRNA for active translation (McCullen et al., 2010). By contrast, the negative regulator OxyS sRNA is thought to repress *rpoS* translation by titrating Hfq because no complementarity between OxyS and *rpoS* has been detected (Hussein and Lim, 2011; Moon and Gottesman, 2011). Moreover, Hfq may restructure the coding region of the inhibitory stem in the *rpoS* leader to permit efficient translation (Fig. 1.9) (Hammerle et al., 2013).

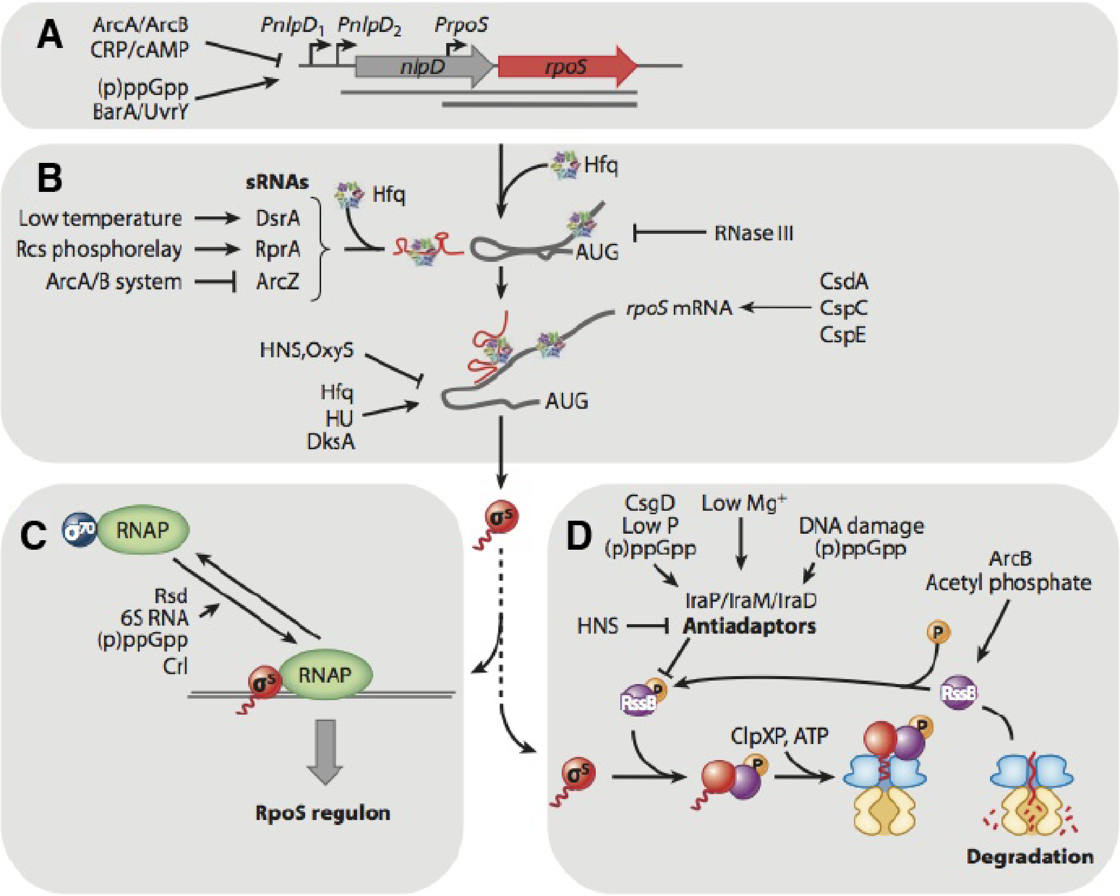


Figure 1.8. Regulation of RpoS expression (reprinted from "The RpoS-mediated general stress response in Escherichia coli") (Battesti et al., 2011). Expression of the alternative transcription factor RpoS (s^S) is regulated at the levels of (*A*) transcription, (*B*) translation, and (*D*) post-translation. (*C*) In stationary phase and under stress conditions, RpoS regulates downstream gene expression by associating with free core RNA polymerase.

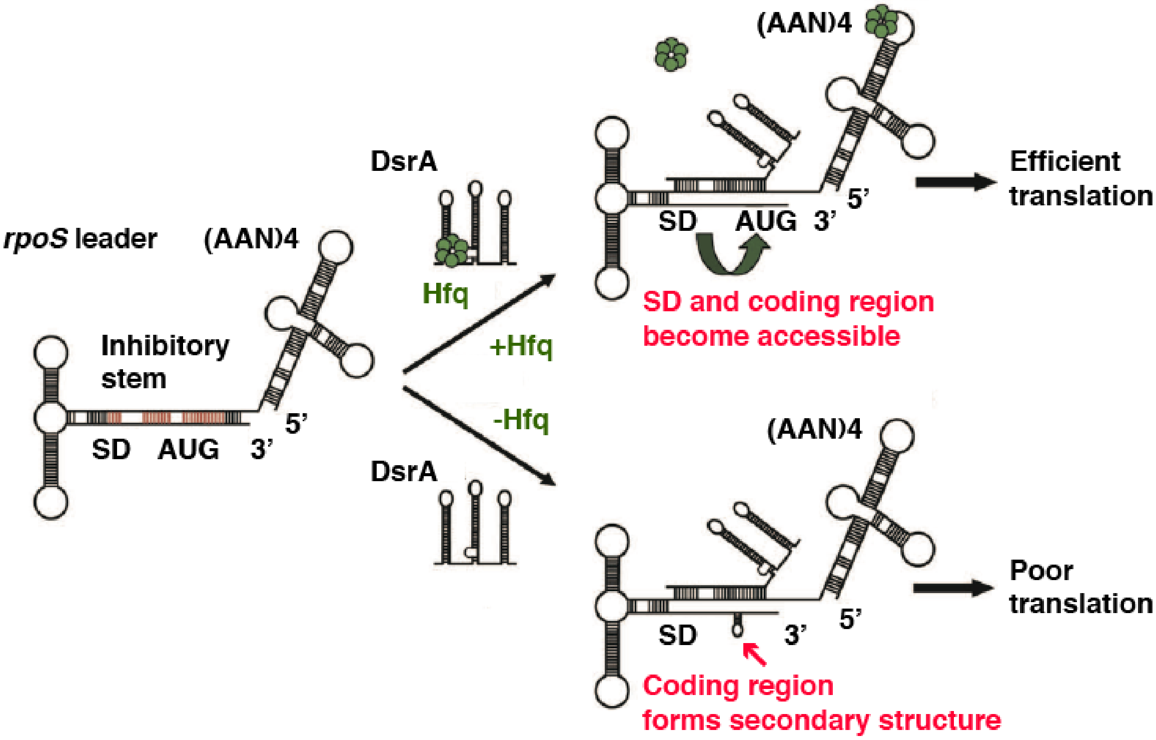


Figure 1.9. Hfq regulates translation initiation of *rpoS* (adapted from "Duplex formation between the sRNA DsrA and rpoS mRNA is not sufficient for efficient RpoS synthesis at low temperature") (Hammerle et al., 2013). *rpoS* leader forms inhibitory stem that contains the sRNA annealing site base paired with and the Shine-Dalgarno sequence (SD) and the coding region. Upper scheme: Hfq recruits DsrA to the sRNA annealing site and unwinds the secondary structure of the coding region in the *rpoS* leader. Once base pairing is complete, Hfq cycles off DsrA and remains bound to the upstream $(AAN)_4$ motif. Lower scheme: in the absence of Hfq, DsrA anneals with the *rpoS* leader to resolve the inhibitory stem; however, local secondary structure within the coding region reduces translation efficiency.

Hfq facilitates sRNA-mRNA annealing

The example of *rpoS* demonstrates one well-established function of Hfq, which is facilitating sRNA-mRNA annealing (Fig. 1.7) (Kawamoto et al., 2006; Lease and Woodson, 2004; Moller et al., 2002; Vecerek et al., 2003), conventionally attributed to two mechanisms. Firstly, Hfq "passively" bridges the sRNA and the mRNA by simultaneously binding to the two strands (Mikulecky et al., 2004). This model proposes that Hfq increases the transient concurrence of RNA strands and brings the complementary sequences into proximity (Storz et al., 2004). In fact, *rpoS* and *flhA* mRNA leaders can form a stable ternary complex with the sRNAs and Hfq (Salim and Feig, 2010; Soper and Woodson, 2008) only when sufficient complementarity exists between the annealing RNAs (Salim and Feig, 2010; Soper et al., 2011).

The second mechanism postulates Hfq as an RNA chaperone that "actively" alters the RNA secondary structure. The first direct evidence dated back to *in vitro* study in 1997, when the protein was found to melt the 3'-end of the Qβ RNA template for replicase access (Schuppli et al., 1997). Subsequent RNase footprinting data confirmed that Hfq significantly changed RNA structures in the regions that anneal with other RNAs including *rpoS* mRNA (Brescia et al., 2003; Geissmann and Touati, 2004; Lease and Woodson, 2004; Moller et al., 2002; Zhang et al., 2002).

Hfq interacts with other protein factors

In addition to facilitating sRNA-mRNA interactions, Hfq binding also affects accessibility of other proteins or complexes that bind the same RNAs. For instance, Hfq directly competes with ribosome binding to the *ompA* mRNA leader and causes translation repression and mRNA degradation (Vytvytska et al., 2000). In the same line, Hfq was recruited by Spot 42 sRNA to the *sdhC* mRNA to block ribosome entry (Desnoyers and Masse, 2012). This "blocking" effect by Hfq also protects some RNAs from RNase E degradation, because RNase E cleavage site coincides with the A/U-rich Hfq binding site (Fig. 1.10A) (Moll et al., 2003).

On the flip side of competing for RNA binding, Hfq was found to associate with 30 proteins or protein-complexes either directly or via nucleic acid interaction (Butland et al., 2005). The best-characterized example is the direct binding between Hfq and RNase E, which induces the degradation of various mRNAs and sRNAs (Fig. 1.10B) (Davis and Waldor, 2007; Ikeda et al., 2011; Morita et al., 2005; Viegas et al., 2007; Zhang and Hong, 2009). One study suggested that Hfq interacts with ribosomal protein S1 in the presence of RNAP (Sukhodolets and Garges, 2003), raising the possibility that Hfq may assist the translation elongation step (Le Derout et al., 2010). Additionally, Hfq indirectly associates with protein complexes to regulate RNA metabolism. For instance, Hfq stimulates poly(A) polymerase activity at the RNA tails (Fig. 1.10C) (Folichon et al., 2005). In another example, Hfq inhibits transcription antitermination at Rho-dependent terminators by regulating Rho activities (Rabhi et al., 2011). More importantly, Hfq was reported to cooperate with global regulatory proteins RsmA and CspC in *Pseudomonas aeruginosa* and *Bacillus subtilis* (Cohen-Or et al., 2010; Sorger-Domenigg et al., 2007).

Clearly, increasing evidence of the canonical and noncanonical Hfq function is beginning to unveil an extensive regulatory network stretching far beyond a humble RNA-binding protein. Nonetheless, at the core of the network lies the fundamental question: how does Hfq recognize and interact with such an astonishing variety of RNA targets? The author has been inspired to pursue this thesis study to address some of the key questions in the Hfq research field.

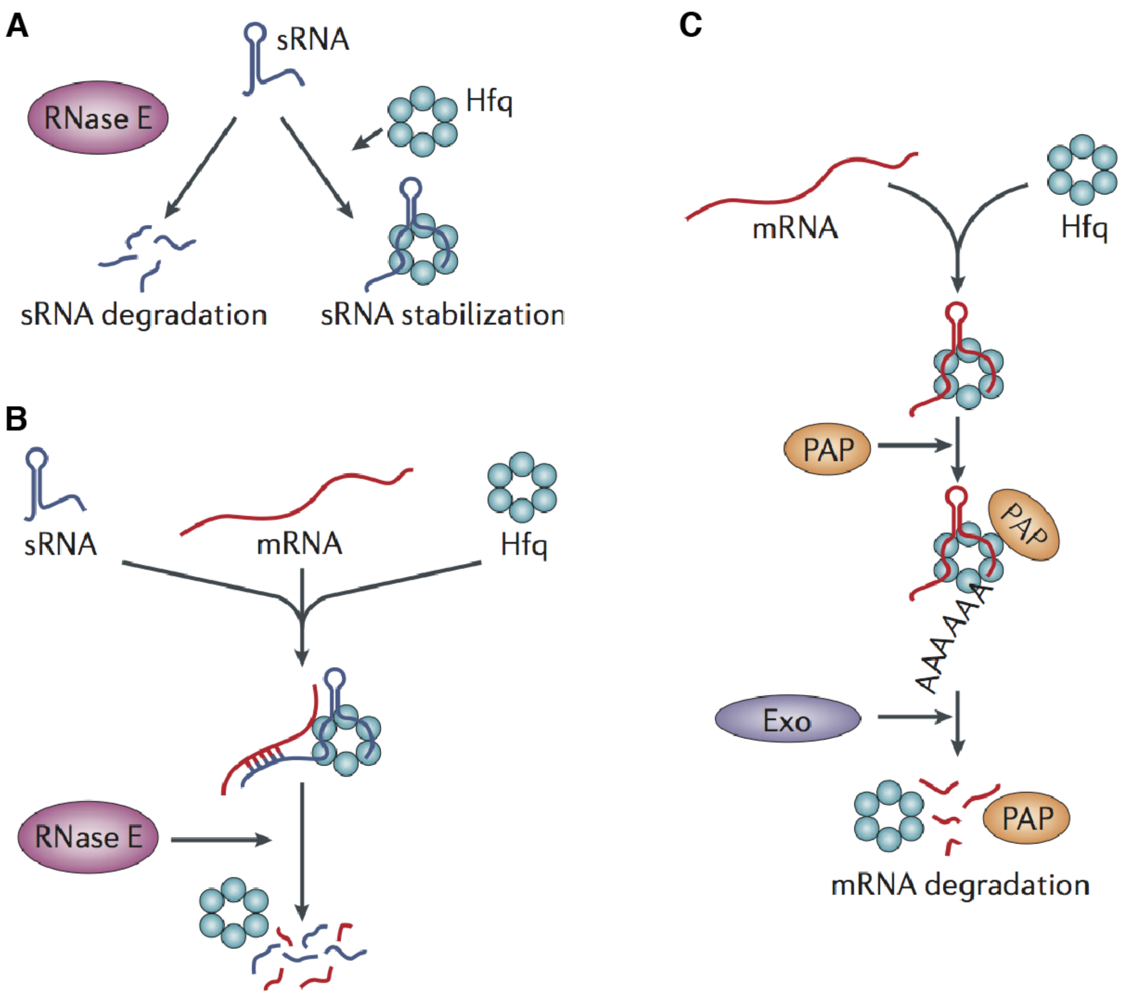


Figure 1.10. Hfq interacts with other protein factors to regulate mRNA metabolism (adapted from "Hfq and its constellation of RNA") (Vogel and Luisi, 2011). (*A*) Hfq protects RNAs from degradation by blocking RNase E binding site. (*B*) Hfq induces RNA degradation by recruiting RNase E. (*C*) Hfq stimulates the polyadenylation of an mRNA by poly(A) polymerase (PAP), leading to 3'-to-5' degradation by an exoribonuclease (Exo).

Thesis outline

During my three-year study in Dr. Sarah Woodson's laboratory, the Hfq community embraced significant advances in the understanding of Hfq•RNA interactions, and also encountered unprecedented challenges to revise the existing Hfq•RNA binding models. The first milestone was the identification of Hfq's lateral rim as a third RNA binding site (Panja et al., 2013; Sauer and Weichenrieder, 2011; Sauer et al., 2012). This discovery triggered many intriguing questions, such as which RNAs bind to the lateral rim, the function of this binding site, and how it relates to the distal and proximal binding surfaces. Secondly, low-resolution molecular envelopes of Hfq•sRNA complexes showed that Hfq partially unwound and altered sRNA structures (Henderson et al., 2013), leading to a series of questions to address: whether Hfq also alters other RNA structures, how Hfq affects RNA tertiary conformation, and how RNA conformational change relates to Hfq's activity. Finally, an Hfq-binding signature has been proposed to classify the sRNA targets as proximal face dependent and rim/distal face dependent (Zhang et al., 2013), suggesting that highly diversified Hfq regulation may effectively depend on a few defined binding modes.

Meanwhile, my colleagues in the Woodson group made remarkable progress in understanding Hfq's chaperone activity, which set off a great starting point for my research. For instance, Dr. Subrata Panja discovered that close proximity and 3'-orientation of Hfq binding site to strand annealing region enhanced Hfq activity for the oligonucleotides (Panja and Woodson, 2012), which inspired me to test the observation in the *rpoS* mRNA leader. Following Dr. Toby Soper's work of identifying upstream

(AAN) motif as a tight Hfq binding site (Soper et al., 2010; Soper and Woodson, 2008), I started seeking for additional weak binding sites in the *rpoS* leader, which led to further studies of tertiary structure and binding modes for *rpoS* and other RNAs.

In the following chapters, I presented my research on four topics. Chapter 2 showed the positional effects of an ectopic A_{18} motif in *rpoS* regulation by sRNAs and Hfq, suggesting that the natural Hfq binding site is optimal for positive regulation because it recruits Hfq to the mRNA and allows it to act on incoming sRNAs without distorting *rpoS* structure. Chapter 3 compared Hfq binding modes of *rpoS* and *flhD* mRNA leaders, showing that Hfq significantly remodeled *rpoS* secondary structure by simultaneously binding to the A-rich and the U-rich motifs at the distal and the lateral surfaces while *flhD* RNA structure remained unchanged due to the lack of a U-rich motif to bind Hfq. Chapter 4 explored the rigid body and molecular dynamic modeling of *rpoS*•Hfq complex, indicating that Hfq not only folded *rpoS* leader but also positioned the complex for sRNA entry and translation initiation. Chapter 5 investigated Hfq binding profiles and mRNA annealing rates in four sRNAs, suggesting that Hfq binding mode depended on the presence of the A-rich motif while annealing behavior relied on complementarity of the mRNA-sRNA pairs. Finally, Chapter 6 presented a more comprehensive model of Hfq's chaperoning mechanisms, which summarized the collective knowledge of the Hfq community, the Woodson group, and myself.

Chapter 2 Positional effects of the AAN motif in *rpoS* regulation by sRNAs and Hfq

Introduction

Different surfaces of Hfq interact with different RNA sequences. For instance, Hfq recognizes a $(AAN)_4$ motif upstream of the sRNA target site in *rpoS* leader to facilitate annealing of sRNAs to the *rpoS* mRNA (Soper and Woodson, 2008; Updegrove et al., 2008). These AAN motifs specifically bind the distal face of Hfq, and are frequently found in mRNA targets (Link et al., 2009; Mikulecky et al., 2004). By contrast, the lateral rim and proximal face of Hfq interact with U-rich sequences present in the body and 3' tail of many sRNAs on (Sauer, 2013; Schumacher et al., 2002). These distinct sRNA and mRNA binding surfaces on each face of Hfq position their complementary regions to interact with a conserved arginine patch on the rim that acts as the active site for RNA annealing (Panja et al., 2013). This arginine patch was proposed to orient sRNAs for base pairing (Sauer et al., 2012) and is needed to catalyze the formation and release of RNA base pairs (Panja et al., 2013).

Despite our knowledge of RNA binding surfaces of Hfq, the molecular "rules" of Hfq binding sites for optimal regulation efficiency are not clear. *In vitro* experiments on short, unstructured RNA oligonucleotides showed that Hfq anneals two RNAs most efficiently when bound less than 20 nucleotides away from the target, preferably on the 3' side (Panja and Woodson, 2012). This proximity requirement is partially overcome by RNA secondary structure that brings Hfq closer to the target.

Intriguingly, Hfq binding sites in bacterial mRNAs vary widely in their proximity to the sRNA binding site, and are found both upstream and downstream of the sRNA. Hfq binds *shiA* and *sodB* mRNAs <20 nucleotides away from the 3' and 5' of the sRNA annealing site, respectively (Masse and Gottesman, 2002; Prevost et al., 2007). Spot 42 targets were also found to have AAN motifs within 14 nt of the target site (Beisel et al., 2012). By contrast, *flhA* mRNA simultaneously binds Hfq distal and proximal surfaces at a distance of 50 nt and 30 nt to the OxyS sRNA binding site, respectively (Salim and Feig, 2010). The AAN motif in *rpoS* mRNA is ≥ 60 nt upstream of the sRNA binding site. These examples suggest that sequence elements within each mRNA may fine tune interactions between Hfq and the sRNA target region, potentially influencing the regulatory outcome.

Here I show that the location of Hfq binding is important for up-regulation of *rpoS* expression by sRNAs in *E. coli*. Using an improved model of the *rpoS* mRNA secondary structure as a guide, I deleted the natural $(AAN)_4$ and A_6 Hfq binding motifs and introduced artificial A_{18} binding sites at different positions upstream and downstream of the sRNA target site. Though *in vitro* assays showed that Hfq binds A_{18} sequences anywhere in the *rpoS* mRNA, Hfq only supports sRNA regulation *in vivo* when recruited to its original location or a site immediately upstream of the sRNA target site. I propose that the native Hfq binding site is optimal because it can communicate with the downstream inhibitory stem-loop through a flexible hinge between two structural domains of the *rpoS* leader.

Results

The full-length *rpoS* leader forms three domains

A previous model of a 323 nt fragment of the *rpoS* mRNA showed that the single-stranded $(AAN)_4$ Hfq binding motif was imbedded in secondary structure upstream of the inhibitory stem-loop (Soper and Woodson, 2008). To better predict how upstream sequences influence *rpoS* translation, I used the full-length *rpoS* leader starting from the *rpoS* promoter 567 bp upstream of the start codon and ending 30 nucleotides after the start codon. Toby Soper generated the secondary structure model for the entire *rpoS* leader by using SHAPE (Fig. 2.1) (Peng et al., 2014).

As expected, the secondary structure of the full-length leader was consistent with the regulatory elements identified previously. First, the 3' end of the leader formed an inhibitory stem-loop between the sRNA annealing site and the ribosome binding site (Fig. 2.1), identical to the structure based on genetic data (Majdalani et al., 1998) and structure probing of the truncated leader (Lease and Woodson, 2004; Soper and Woodson, 2008). Second, the upstream $(AAN)_4$ and A_6 motifs that bind the distal face of Hfq were again predicted to reside in two single-stranded regions of the leader, separated by a short helix similar to the previous model (Soper and Woodson, 2008).

On the other hand, the new model predicted three additional features likely to be important for translational control (Fig. 2.1). First, the inhibitory stem was extended by 15 bp due to the base pairing between the upstream of the sRNA annealing site and the coding sequence, which may account for the lower basal expression of *rpoS* mRNAs containing the full-length leader (Soper et al., 2010).

Second, the hinge region between the inhibitory stem and the upstream AAN motifs was shortened and joined to the far upstream region as a four-way-junction. This region is dynamic and could allow for tertiary contacts between the upstream and downstream domains.

Third, the far upstream region was predicted to form four extended helices, indicating that the 5’ half of the leader forms a highly stable secondary structure. Thus, the full-length *rpoS* leader can be divided into three domains: the far upstream domain, the A-rich domain containing the strong Hfq binding site, and the downstream inhibitory stem domain.

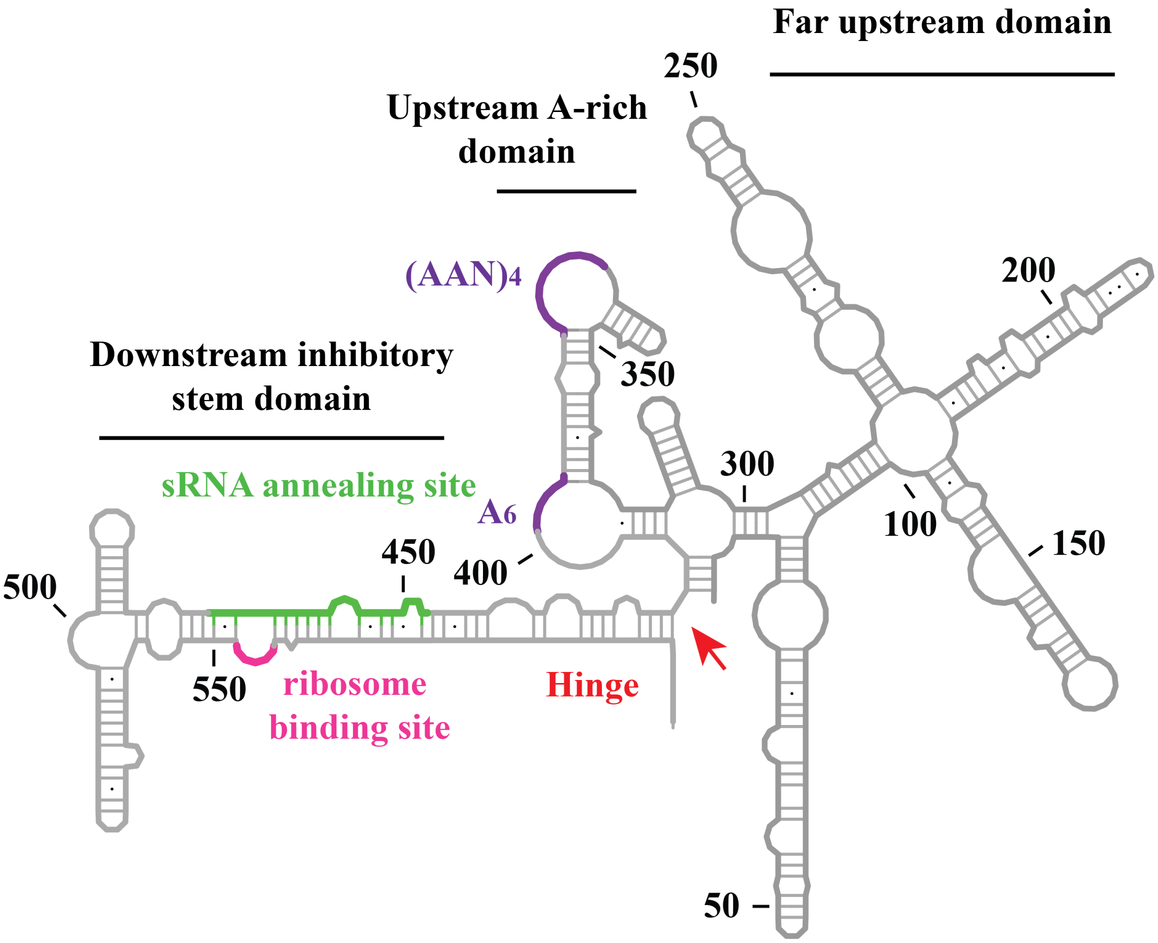


Figure 2.1. Structural model of the full-length *rpoS* leader (adapted from "Positional effects of the AAN motif in *rpoS* regulation by sRNAs and Hfq" (Peng et al., 2014). Full-length *rpoS* leader contains the far upstream domain, the upstream A-rich domain and the downstream inhibitory stem connected by a flexible hinge. Important functional elements are labeled accordingly.

Ectopic Hfq binding sites rescue *rpoS* regulation in vivo

I next attempted to understand how the functional elements of the *rpoS* leader were organized to interact with Hfq, the regulatory sRNAs, and the translation machinery. Upstream $(AAN)_4$ and A_6 motifs specifically bind the distal face of Hfq and contribute to sRNA regulation (Soper et al., 2010; Soper and Woodson, 2008). I asked whether the position of the Hfq binding site is important for regulation by relocating A-rich motifs to other regions of the *rpoS* leader.

A double mutant (*rpoSΔ2*) lacking both the $(AAN)_4$ and A_6 motifs cannot bind Hfq tightly, and consequently Hfq no longer facilitates its annealing with the sRNAs (Soper et al., 2010). Using the secondary structure model as a guide, I inserted an A_{18} sequence cassette at various positions in the three domains of *rpoSΔ2* (Fig. 2.2A). Insertion sites chosen for further study were in the far upstream region (positions 53 and 250), the original $(AAN)_4$ loop (position 366), 5’ of the inhibitory stem (position 441), and the 3’ of the sRNA binding site (positions 484, 499, and 519) (Fig. 2.2A). All of these sites were designed to minimize perturbations to the mRNA structure, as predicted by MFOLD (Zuker, 2003). Partial SHAPE modification data of Δ2-366A18 and Δ2-519A18 RNAs indicated these insertions do not grossly change the secondary structure of the inhibitory domain (data not shown).

To determine whether these ectopic Hfq binding sites could support *rpoS* regulation *in vivo* (Fig. 2.2B), the *rpoS* leaders were fused to the *lacZ* reading frame in the *E. coli* chromosome as previously described (Mandin and Gottesman, 2009). RpoS translation was induced by overexpression of DsrA, RprA, or ArcZ sRNA (pLac), and expression of the *rpoS-lacZ* fusion was obtained from the specific β-galactosidase activity of cell cultures in the late-log to early-stationary phase.

As expected, *rpoSΔ2-lacZ* fusions lacking the upstream AAN motifs were translated 30% - 50% less than the wild type *rpoS-lacZ* fusion, when DsrA, RprA or ArcZ sRNAs were over-expressed in the presence of IPTG (Fig. 2.2C, green, yellow and red bars). Basal expression of *rpoSΔ2-lacZ* in the presence of empty vector (pLac) was also 30% lower that for the wild type *rpoS* fusion (Fig. 2.2C, blue bars), indicating that the AAN motifs are important for activation by endogenous sRNAs. As cells over-expressing DsrA (Fig. 2.2C, green bars) can induce *rpoS* without Hfq (Soper et al., 2010), these A_{18} strains were least affected by deletion of the AAN motifs.

I next asked whether A_{18} insertions could rescue Hfq-dependent translation of *rpoSΔ2*. When the A_{18} Hfq binding site was located far upstream in the *rpoS* leader (positions 53 and 250), the expression levels remained the same, suggesting these sites are too far from the sRNA target site. When A_{18} was inserted at the natural location of the $(AAN)_4$ motif (position 366) or 5' of the sRNA target site (position 441), basal and activated expression improved about 2-fold compared to *rpoSΔ2* (Fig. 2.2C, Δ2-366A18 and Δ2-441A18), suggesting that activation by endogenous and over-produced sRNAs is restored when Hfq binds near the sRNA target site.

By contrast, A_{18} insertions 3' of the sRNA target site did not rescue *rpoS* expression (positions 484, 499, and 519). In fact, insertions at positions 484 and 499 decreased the level of expression by half under some conditions (Fig. 2.2C, Δ2-484A18 and Δ2-499A18). Therefore, the level of expression depended on the position of the A_{18}

insertion and was optimal when this sequence was placed just upstream of the sRNA (Fig. 2.2C).

rpoS regulation requires 5’ Hfq binding *in vivo*

If the *rpoS*Δ2-A18 fusions rescue expression by recruiting Hfq to the mRNA, then deleting *hfq* from the cells should erase any change of expression induced by A_{18} insertions. I deleted *hfq* from strains containing the *rpoS::lacZ* fusions, and re-measured the ability of sRNAs to activate *rpoS* translation. Deleting *hfq* lowered expression of all the *rpoS* fusions tested (Fig. 2.2D), apart from strains over-expressing DsrA that do not require Hfq to induce *rpoS* expression (Soper et al., 2010). Moreover, the Δ2-366A18 and Δ2-484A18 fusions were expressed at a level similar to the *rpoSΔ2* mutant in the *hfq*$^-$ strains. This indicates that the different activities of the A18-containing leaders arise from their ability to bind Hfq, and not from a difference in the mRNA stability or translation levels.

When A_{18} was inserted at position 441 immediately 5’ of the sRNA binding site, expression was slightly higher even without Hfq. I suspect that this insertion weakens the inhibitory stem-loop, causing leaky repression of translation initiation. In fact, none of the *rpoSΔ2*-A18 reporters were as active as the wt *rpoS*::lacZ fusion, possibly because the $(AAN)_4$ deletion alters the structure of the Hfq binding domain (Soper et al., 2010). Nonetheless, the ability of A_{18} insertions to partially restore *rpoS* translation in the presence of Hfq suggested that sRNA regulation requires recruitment of Hfq to an AAN binding site within 80 nt upstream of the sRNA.

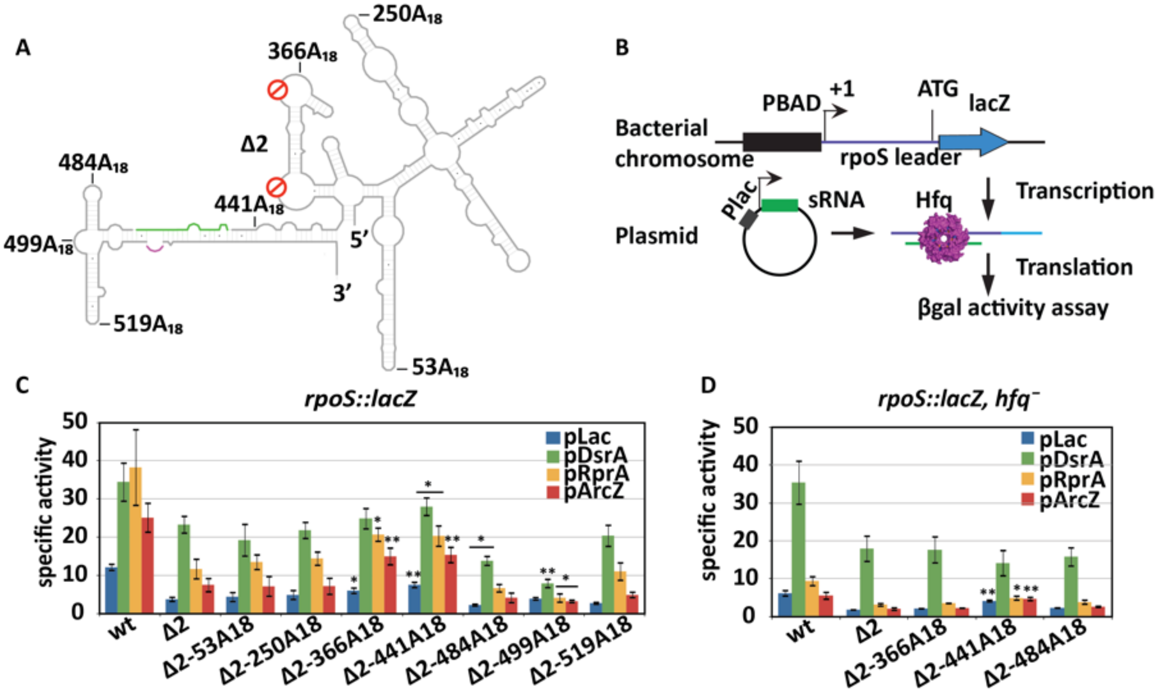


Figure 2.2. Hfq location determines *rpoS* translational activation *in vivo*. (*A*) Schematic representation of A_{18} insertions in the full-length *rpoS* leader lacking both A-rich motifs (Δ2). The positions are numbered from the *rpoS* transcription start site. The sRNA annealing site, green; Shine-Dalgarno, purple. (*B*) Experimental design. The *rpoS* leader sequence (606 nt) was fused with a chromosomal *lacZ* downstream of a P_{BAD} promoter. DsrA, RprA, or ArcZ sRNA was overexpressed from the P_{lac} promoter on plasmids. Hfq was expressed from its endogenous gene. (*C*) sRNA activation of *rpoS-lacZ* translation. Specific activity of b-galactosidase in strains carrying the *rpoS-lacZ* fusion listed below the x-axis, and transformed with the control plasmid pLac (blue), or plasmids overexpressing DsrA (green), RprA (yellow), and ArcZ (red). Error bars represent the standard deviation of three independent trials. Strains carrying Δ2-A18-*lacZ* fusions were compared to the one carrying Δ2-*lacZ* fusions for their specific activities by unpaired one-tail T-test. Significantly different expression levels are marked with * ($p < 0.05$) or ** ($p < 0.01$). Expression levels of all the mutants were significantly lower than wt with $p < 0.01$. (*D*) sRNA activation of *rpoS-lacZ* fusions in an *hfq*$^-$ background. Same as in C. Mutant *rpoS-lacZ* fusions were compared to wt *rpoS-lacZ* fusion by unpaired one-tail T-test; * ($p < 0.05$) or ** ($p < 0.01$).

A_{18} insertion restores tight Hfq binding in vitro

I next analyzed how the A_{18} insertions at different positions affect Hfq binding to the *rpoS* mRNA leader. Based on the structural model of the full-length leader (Fig. 2.1), I designed a new 301 nt *rpoS* leader that lacks the far upstream helices (nt 9–300) that are not required for Hfq binding (Soper and Woodson, 2008), but retains the 5’ fragment (nt 1–8) that forms part of the hinge region. The resulting *rpoS301* leader contained the upstream $(AAN)_4$ and A_6 motifs, the flexible hinge, and the downstream inhibitory stem (Fig. 2.3A). It migrated as single species in native polyacrylamide gel, indicating more uniform folding than the previous *rpoS323* truncation (Soper et al., 2010; Soper and Woodson, 2008).

Soper et al. previously found that a 323 nt *rpoS* RNA bound multiple Hfq hexamers in native (TBE) polyacrylamide gels, described by one tight and several nonspecific binding sites with dissociation constants $K_T = 0.28$ µM and $K_{nsp} = 1$ µM Hfq monomer (Soper and Woodson, 2008). To better discriminate specific from non-specific Hfq binding, I supplemented the running buffer with 2 mM Mg^{2+} to stabilize the folded RNA. Under these conditions, I observed two specific *rpoS*•Hfq complexes (R•H and R•H2) for WT *rpoS323* (data not shown) and *rpoS301* (Fig. 2.3B) with dissociation constants $K_1 = 0.30$ µM and $K_2 = 0.42$ µM Hfq monomer corresponding to distinct transitions in the Hfq titrations (Fig. 2.3C). I also included a nonspecific binding term in the partition function ($K_{nsp} = 0.87$ µM) to account for high molecular weight complexes that did not enter the gel.

As expected, the *rpoS301Δ2* RNA formed only one specific RNP complex (Fig. 2.3D) with $K_2 = 0.40$ µM Hfq monomer (Fig. 2.3E), which corresponds to weak Hfq binding. Therefore, the Δ2 mutation abolished the strongest Hfq binding site, which presumably corresponded to the $(AAN)_4$ and A_6 motifs (K_1 ~ 0.3 μM Hfq monomer) (Soper and Woodson, 2008). The second binding site, K_2 ~ 0.4 μM, also behaved like a specific binding interaction, because it generated a second stable RNP and produced very similar K_d values for wild type and Δ2 *rpoS301*. By contrast, non-specific binding produced high molecular weight complexes that were poorly resolved in these gels. Because low affinity non-specific complexes (0.7–1 μM Hfq) were treated collectively in my partition function, the resulting K_{nsp} values varied considerably among different *rpoS* RNAs.

With my optimized *rpoS301* truncation and improved Hfq titration protocol, I measured Hfq binding constants for all the *rpoSΔ2*-A18 fusions. All the A_{18} insertions restored tight Hfq binding with K_1 values comparable to the wild type *rpoS301* (0.26 to 0.33 μM; Fig. 2.3F, red bars). Like WT *rpoS*, *rpoSΔ2*-366A18, Δ2-441A18, and Δ2-519A18 formed a second specific complex with K_2 ~ 0.4 μM (Fig. 2.3F, blue bars, and Fig. 2.4A, B and E). Surprisingly, the downstream insertions Δ2-484A18, and Δ2-499A18 (Fig. 2.4C and D) formed the first Hfq complex but not the second complex (Fig. 2.3F, blue bars). These insertions may disrupt the second binding site, or indirectly disrupt the second RNP by changing the *rpoS* mRNA conformation. Taken together, all the A_{18} insertions tested were capable of binding Hfq, showing that Hfq recognition is independent of context in the RNAs. However, insertions at positions 484 and 499 in the inhibitory stem-loop domain disrupted a second specific interaction with Hfq.

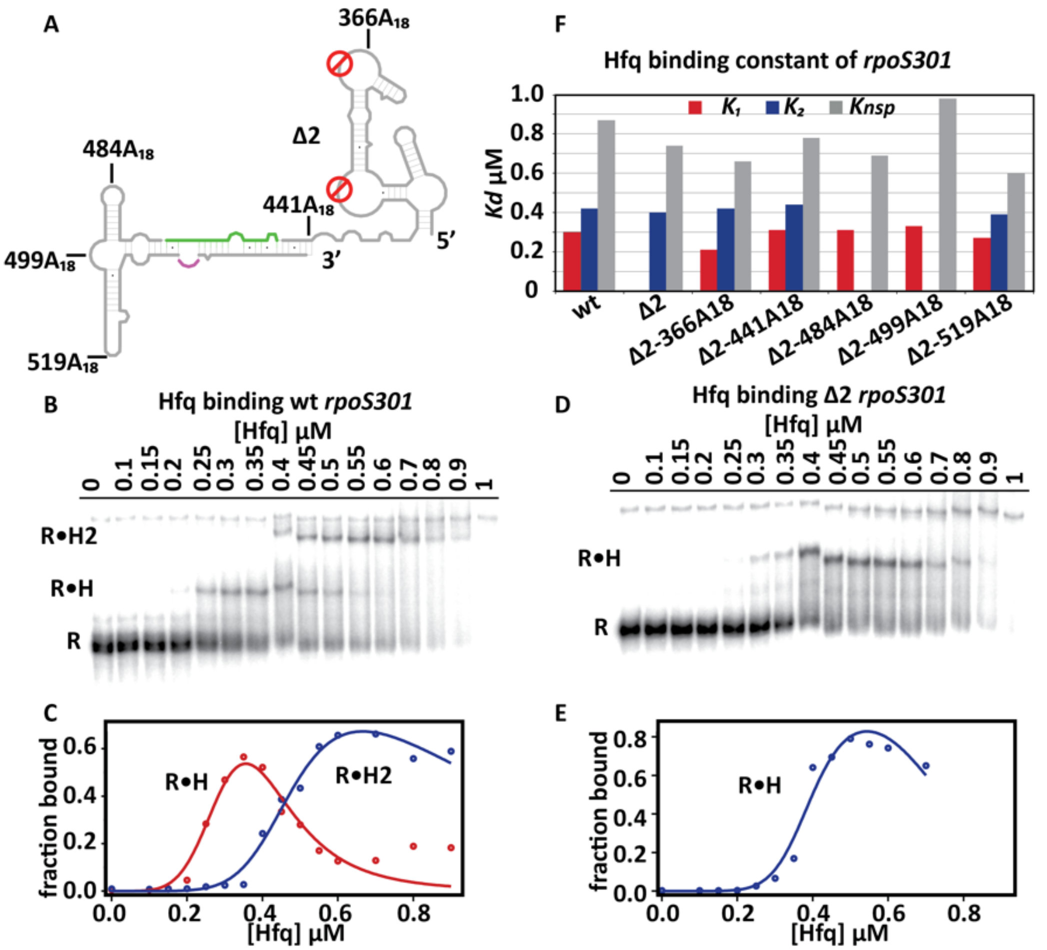


Figure 2.3. Hfq binds *rpoS* mRNA with A_{18} insertions. (*A*) Schematic representation of the *rpoS301* leader that lacks the far upstream domain. (*B*) Hfq titrations of uniformly labeled wt *rpoS301* at 25ºC. Free *rpoS* (R) binds one or two Hfq hexamers (R•H and R•H2). A smear of complexes at the top of the gel in high Hfq are not shown. Around 1% of the *rpoS*301 formed a slow migrating band that did not bind Hfq. (*C*) Fraction of bound wt *rpoS301* as a function of [Hfq] was fit to eq. 2.2 (Methods). Red, R•H (k_1); blue, R•H2 (k_2). The fraction of R•H and R•H2 decreased at high [Hfq] due to formation of high molecular weight (non-specific) complexes. (*D,E*) Hfq titrations of uniformly labeled *rpoS301* Δ2 at 25ºC as in (*B*) and (*C*), except data were fit to eq. 2.1. (*F*) Comparison of the Hfq binding constants among *rpoS301* wt, Δ2, and Δ2-A18 fusions. Data for Δ2-A18 fusions are shown in Fig. 2.4; K_d values are listed in Table 2.1. For *rpoS301* wt, S.D. < 0.02 µM (7%) among 3+ trials.

Figure 2.4. Hfq binding of *rpoS301* Δ2-A18 fusions in vitro. Hfq binding to variants of *rpoS301* RNA was measured by native gel mobility shift as described in Fig. 2.3. (*A*) to (*E*) Δ2-366A18, Δ2-441A18, Δ2-484A18, Δ2-499A18, Δ2-519A18. Left, Hfq titration of *rpoS* resolved on native 6% polyacrylamide gels in THEM$_2$. Right, fraction of bound *rpoS* fit with eq. 2.2. Δ2-366A18 and Δ2-441A18 formed two distinct RNP complexes corresponding to the tight and weak binding sites; Δ2-519A18 formed three RNP complexes corresponding to the tight, weak, and nonspecific binding sites. By contrast, Δ2-441A18 formed only one distinct RNP complex corresponding to the tight binding site. The smear around the R•H2 position may result from the unstable second RNP band. Similarly, Δ2-484A18 formed the first RNP complex and a faint second RNP band at high [Hfq], corresponding to the tight and nonspecific binding sites, respectively. Both binding curves were fit with eq. 2.1. The K_d values are summarized in Table 2.1.

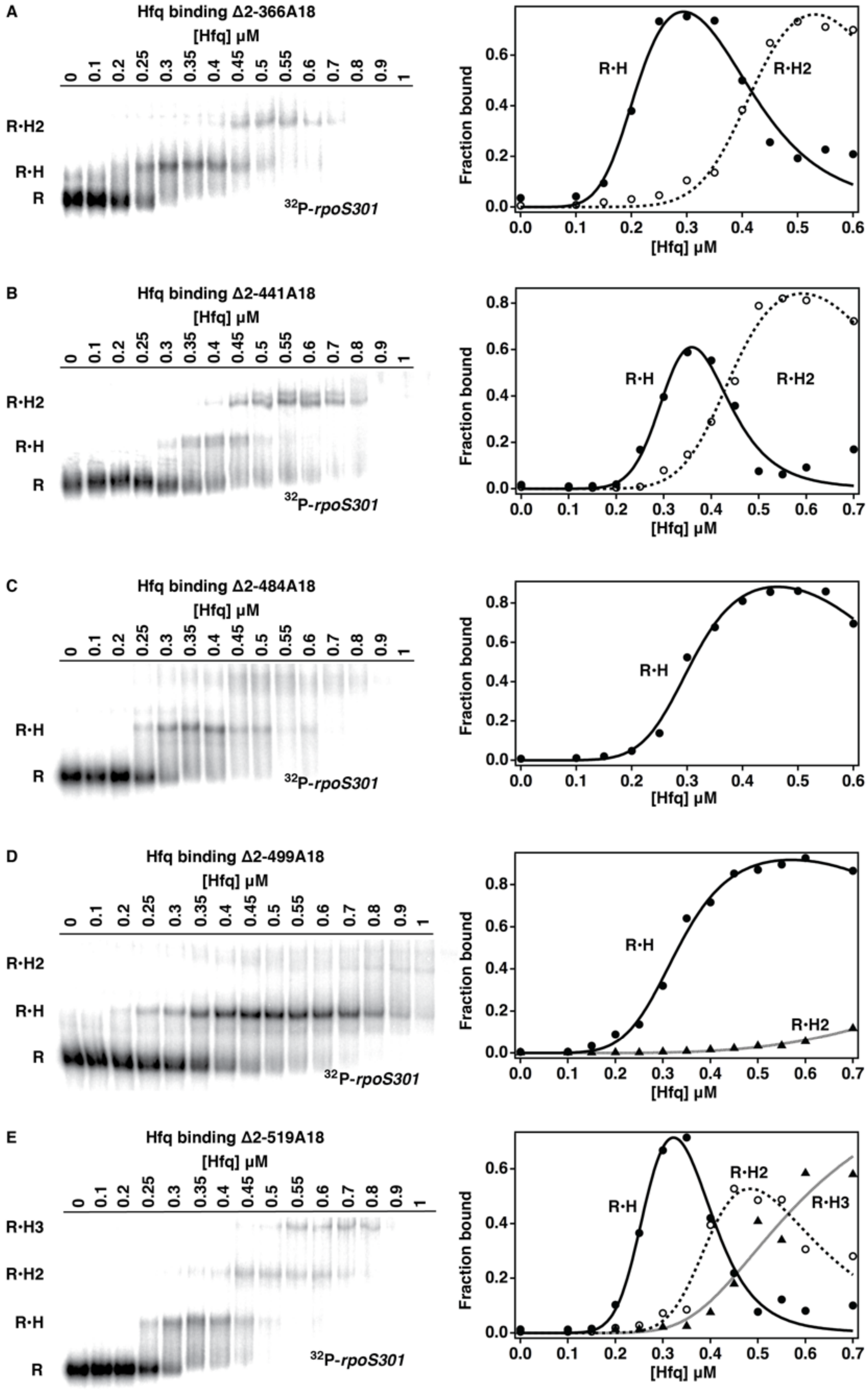

A
Hfq binding Δ2-366A18
[Hfq] μM
0 0.1 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.7 0.8 0.9 1
R·H2
R·H
R
32P-rpoS301
Fraction bound
R·H
R·H2
[Hfq] μM
B
Hfq binding Δ2-441A18
[Hfq] μM
0 0.1 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.7 0.8 0.9 1
R·H2
R·H
R
32P-rpoS301
Fraction bound
R·H
R·H2
[Hfq] μM
C
Hfq binding Δ2-484A18
[Hfq] μM
0 0.1 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.7 0.8 0.9 1
R·H
R
32P-rpoS301
Fraction bound
R·H
[Hfq] μM
D
Hfq binding Δ2-499A18
[Hfq] μM
0 0.1 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.7 0.8 0.9 1
R·H2
R·H
R
32P-rpoS301
Fraction bound
R·H
R·H2
[Hfq] μM
E
Hfq binding Δ2-519A18
[Hfq] μM
0 0.1 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.7 0.8 0.9 1
R·H3
R·H2
R·H
R
32P-rpoS301
Fraction bound
R·H
R·H2
R·H3
[Hfq] μM

Table 2.1 Summary of *rpoS*•Hfq binding constants and *rpoS*•DsrA annealing rate

	Hfq binding constant K_d (μM)			DsrA annealing rate k_{obs} (min^{-1})		
rpoS301	K_1	K_2	K_{nsp}	-Hfq	+Hfq	fold of increase
wt	0.3	0.42	0.87	0.04	2.5	62.50
Δ2	-	0.4	0.74	0.012	0.018	1.50
Δ2-366A18	0.21	0.42	0.66	0.016	1.24	77.50
Δ2-441A18	0.31	0.44	0.78	0.0016	0.095	59.38
Δ2-484A18	0.31	-	0.69	0.011	0.16	14.55
Δ2-499A18	0.33	-	0.98	0.014	0.23	16.43
Δ2-519A18	0.27	0.39	0.6	0.016	0.49	30.63
ΔA6	0.33	0.43	1.04			
Δ(AAN)4	0.26	0.42	1.06			
Δ484	0.46	0.55	1.07			
Δ499	0.44	0.59	1.17			
Δ519	0.41	0.55	1.25			

The binding properties were measured by native polyacrylamide gel electrophoresis, as shown in Fig. 2.3-2.6 and 2.8. The equations and calculations were described in Material and Methods.

5' Hfq binding facilitates DsrA•*rpoS301* annealing in vitro

While Hfq was able to bind A_{18} sequences at any position in the *rpoS* leader, only A_{18} insertions 5' of the sRNA target sequence rescued sRNA regulation *in vivo*. I next determined whether Hfq must be recruited to a specific location in the *rpoS* leader to facilitate annealing of DsrA sRNA with the *rpoS301* RNA using native gel mobility shift assays. As expected, while WT *rpoS301* annealed with DsrA slowly in the absence of Hfq (Fig. 2.5A), it rapidly formed a ternary complex with DsrA and 0.6 μM Hfq (Fig. 2.5B). Consistent with previous studies (Soper et al., 2010), the initial annealing rate with Hfq was 2.5 min^{-1}, 62 times faster than with no Hfq (0.04 min^{-1}; Fig. 2.5C and D). By contrast, *rpoS301*Δ2 was unable to form the ternary complex, and the annealing kinetics was about the same with and without Hfq (0.018 min^{-1} and 0.012 min^{-1}, respectively; Fig. 2.6A).

I then compared the DsrA annealing kinetics with or without Hfq for all the *rpoS301* RNAs containing A_{18} insertions. All the insertions except Δ2-441A18 had little effect on the annealing rate in the absence of Hfq (ranging from 0.011 to 0.016 min^{-1}), suggesting that the DsrA•*rpoS* mRNA interaction alone was not altered by an A_{18} insertion at most positions (Fig. 2.6, left column). The Δ2-441A18 RNA was almost unable to base pair with DsrA without Hfq (Fig. 2.6, C left column), possibly because the insertion at nt 441 changes the *rpoS* mRNA structure (Fig. 2.3A).

All the A_{18} insertions allowed Hfq to facilitate sRNA annealing to some degree, depending on the location. Firstly, unlike *rpoSΔ2*, they all formed stable ternary complexes with Hfq and DsrA (Fig. 2.6, middle column), indicating that any A-rich Hfq binding site in the *rpoS* leader was necessary and sufficient for stable ternary complexes.

Secondly, Hfq accelerated DsrA annealing with all the Δ2-A18 RNAs (Fig. 2.6, right column), but insertions at the 5’ of the sRNA annealing site restored the annealing kinetics to the same level as WT *rpoS*301, while insertions at 3’ side did not (Fig. 2.5D). These data agreed with my *in vivo* results that Hfq promotes sRNA-*rpoS* mRNA interactions most effectively when the AAN motif is placed 5’ of the sRNA annealing site.

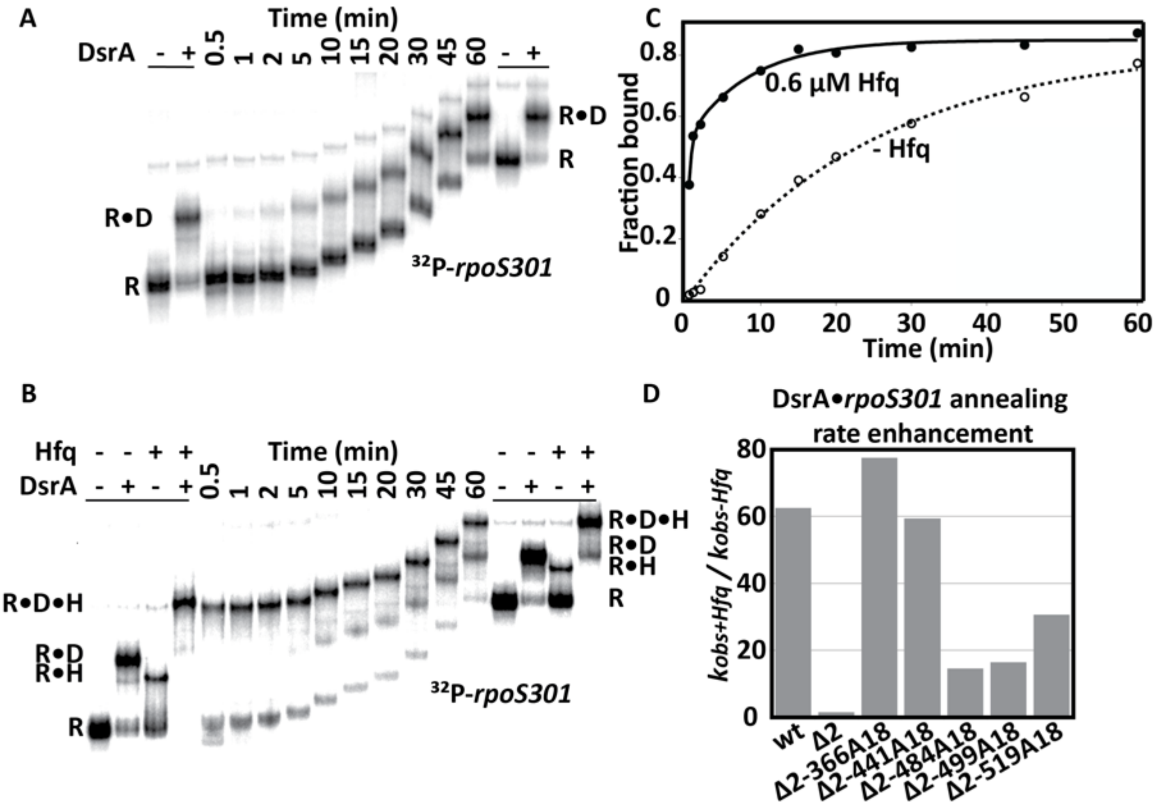
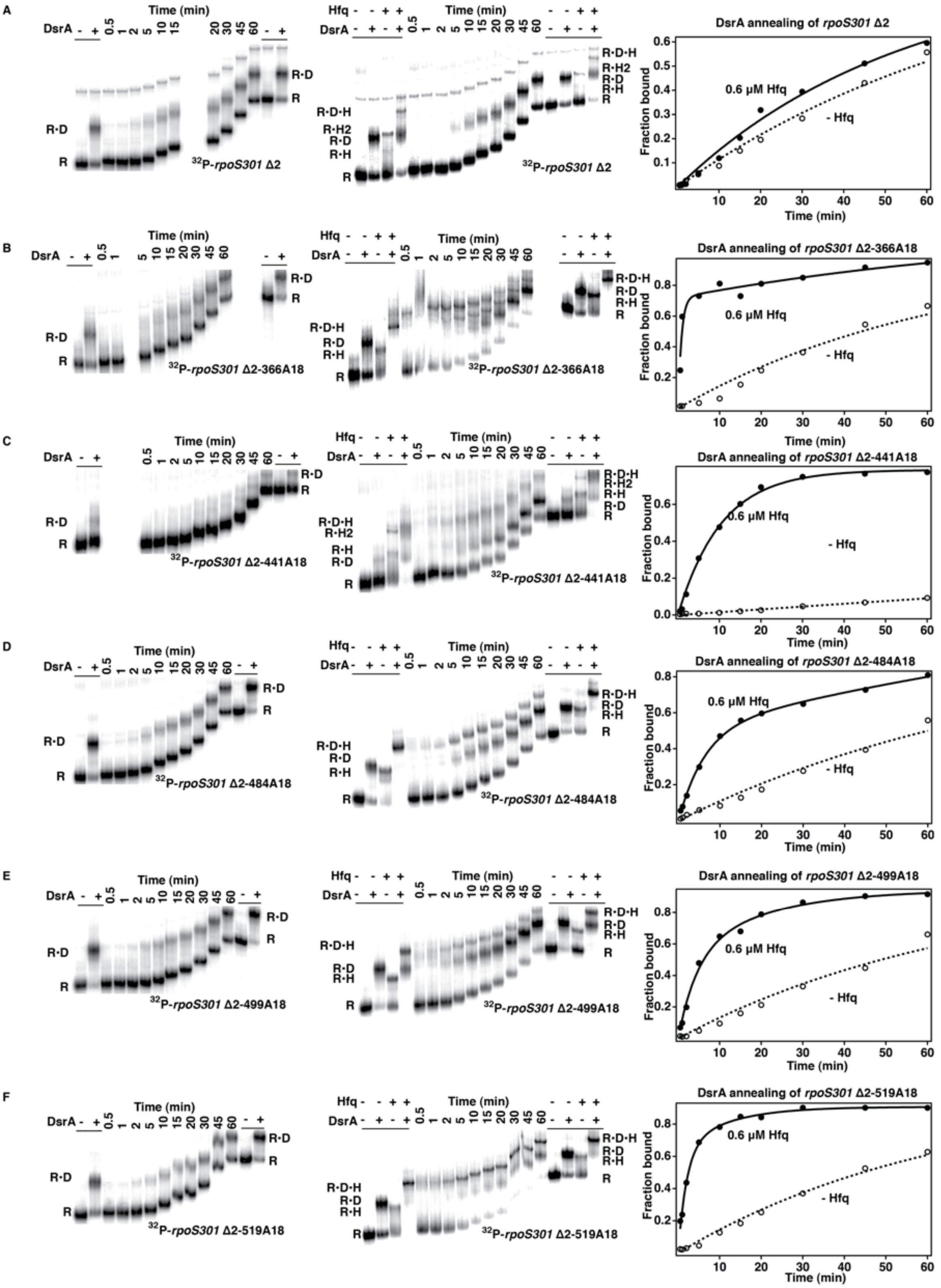


Figure 2.5. Upstream A_{18} insertions rescue sRNA annealing by Hfq. (*A*) 200 nM DsrA was mixed with uniformly ^{32}P-labeled wt *rpoS301* without Hfq at 25ºC 0-60 min. Reactions were loaded on the gel at the times shown above each lane. Markers of R and R•D complexes (+,– lanes) were generated by incubating the same annealing reaction at 25ºC for 2 hr and were loaded before the first time point or after the last time point. (*B*) 200 nM DsrA binding to labeled wt *rpoS301* with 0.6 µM Hfq as in B, with four marker lanes showing R, R•D, R•H, and R•D•H complexes. Free *rpoS301* migrated as multiple bands 0.5 min after the reaction started, which was not observed in B and may reflect a transient conformational change upon addition of Hfq. (*C*) Binding kinetics of *rpoS* and DsrA in the absence (open circles) and presence of Hfq (closed circles). No Hfq, $k_{obs\text{-}Hfq}$ = 0.04 min^{-1} (82.6%); 0.6 µM Hfq, combined R•D and R•D•H complexes formed with a fast phase $k_{obs+Hfq}$ = 2.50 min^{-1} (52 %) and slow phase $k_{obs+Hfq}$ = 0.12 min^{-1} (33%). (*D*) DsrA annealing rates of wt, Δ2, and Δ2-A18 *rpoS301* RNAs reported as the ratio of the fast phase $k_{obs+Hfq}$ and $k_{obs\text{-}Hfq}$. Data for Δ2 and Δ2-A18 insertions are shown in Fig. 2.6; rate constants are listed in Table 2.1.

Figure 2.6. DsrA annealing to *rpoS301* Δ2-A18 RNAs *in vitro*. The annealing kinetics of DsrA and ^{32}P-labeled *rpoS301* RNAs was measured at 25 ˚C as described in Fig. 2.5. Gels were run continuously during the experiment so samples on the right traveled less far than those on the left. (*A*) to (*F*) Δ2, Δ2-366A18, Δ2-441A18, Δ2-484A18, Δ2-499A18, Δ2-519A18. Left, DsrA annealing kinetics without Hfq. Center, DsrA annealing kinetics of *rpoS* with 0.6 µM Hfq monomer. Right, fraction of bound *rpoS* fit with monophasic or biphasic rate equations. Note that with Hfq present, Δ2 formed the R•D•H ternary complex in the marker lane (2 hr) but not over the reaction time course of 60 min, indicating that the ternary complex formation is extremely slow. By contrast, all the Δ2-A18 fusions formed the ternary complex very rapidly. However, the ternary complex collapsed into binary complex faster than for wt *rpoS301* (Fig. 2.5B), suggesting that it is less stable with the mutant RNAs. The k_{obs} values were summarized in Table 2.1. Reactions with no Hfq and for Δ2 with Hfq were fit to eq.2.3; all others were fit to eq. 2.4.


A
B
C
D
E
F
Time (min)
DsrA
Hfq
R·D·H
R·H2
R·D
R·H
R
³²P-rpoS301 Δ2
³²P-rpoS301 Δ2-366A18
³²P-rpoS301 Δ2-441A18
³²P-rpoS301 Δ2-484A18
³²P-rpoS301 Δ2-499A18
³²P-rpoS301 Δ2-519A18
DsrA annealing of rpoS301 Δ2
DsrA annealing of rpoS301 Δ2-366A18
DsrA annealing of rpoS301 Δ2-441A18
DsrA annealing of rpoS301 Δ2-484A18
DsrA annealing of rpoS301 Δ2-499A18
DsrA annealing of rpoS301 Δ2-519A18
0.6 μM Hfq
- Hfq
Fraction bound
Time (min)
0.5 1 2 5 10 15 20 30 45 60

Downstream A-/U-rich sequences stabilize *rpoS* mRNA conformation

So far, A_{18} insertions at 3' of the sRNA annealing sites (positions 484, 499, and 519) were less effective to restore Hfq function *in vivo* and *in vitro*, which is opposite to the conclusion based on oligonucleotide experiments (Panja and Woodson, 2012). I noticed that those positions are each adjacent to a short A-/U-rich patch that might interact directly with Hfq or might form tertiary interactions in the *rpoS* leader. I first tested whether those A-/U-rich sequences are functionally important by converting them to G-/C-rich sequences (Fig. 2.2A; Δ484, Δ499, and Δ519), and measuring the translation activation of those mutants by Hfq and sRNAs based on β-galactosidase assay described above. The Δ484, Δ499, and Δ519 mutations all reduced expression of *rpoS-lacZ* significantly compared to the wild type *rpoS* leader (Fig. 2.7A). Mutation of the U-rich loop at position Δ484 was the most detrimental, decreasing expression levels by average 40%, followed by Δ519 with expression levels decreased by ~30%, and Δ499 with a moderate 20% decrease of expression levels.

I next asked whether these A-/U-rich sequences were Hfq binding sites. The Δ484, Δ499, and Δ519 mutants all formed two specific Hfq complexes in vitro (Fig. 2.8, left column), suggesting that both tight and weak Hfq binding sites were retained even though insertions at these positions disrupted the weak complex. All three single mutations weakened Hfq binding overall, however, increasing K_1 from 0.3 μM to 0.45 μM, and K_2 from 0.4 μM to 0.55 μM (Fig. 2.7B). While I cannot exclude the possibility that the A-/U-rich sequences contact Hfq directly, these data suggested that these loops fold the *rpoS* leader in a manner that favors Hfq recognition.

If my original A_{18} insertions at positions 499 and 519 (Δ2-499A18 and Δ2-519A18) lowered *rpoS* expression by disrupting tertiary interactions formed by the A-/U-rich sequences at those positions, inserting A_{18} into the wild type *rpoS* leader at positions 499 and 519 should cause a similar loss of function. In fact, both wt-499A18 and wt-519A18 showed reduced expression compared to WT *rpoS::lacZ* (Fig. 2.7A). I found that wt-499A18 was more deleterious than wt-519A18 (60 - 64% and 33% - 57% reduction, respectively), consistent with my results with Δ2-499A18 and Δ2-519A18 (Fig. 2.2C). Together, these data suggest that the downstream A-/U-rich sequences are important for *rpoS* function because they stabilize the tertiary conformation of the mRNA and Hfq recognition, and that this accounts for why Hfq binding sites downstream of the sRNA target site reduce *rpoS* up-regulation.

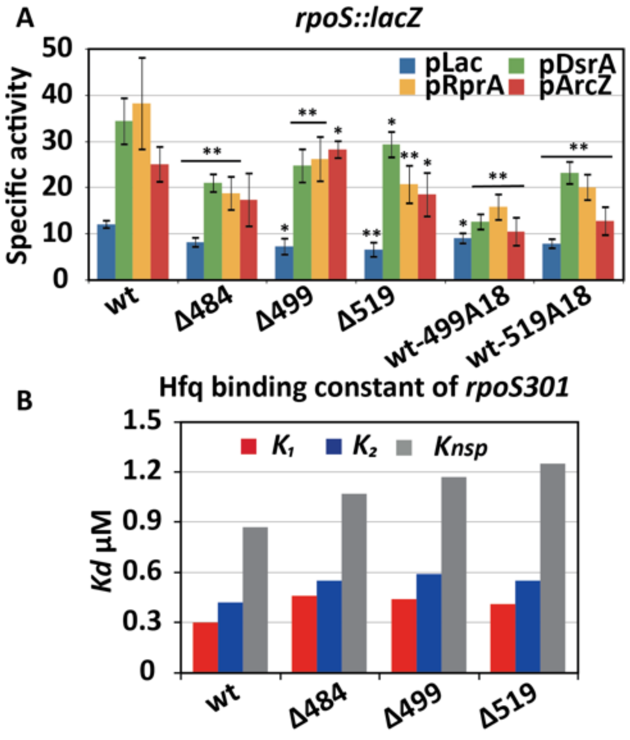
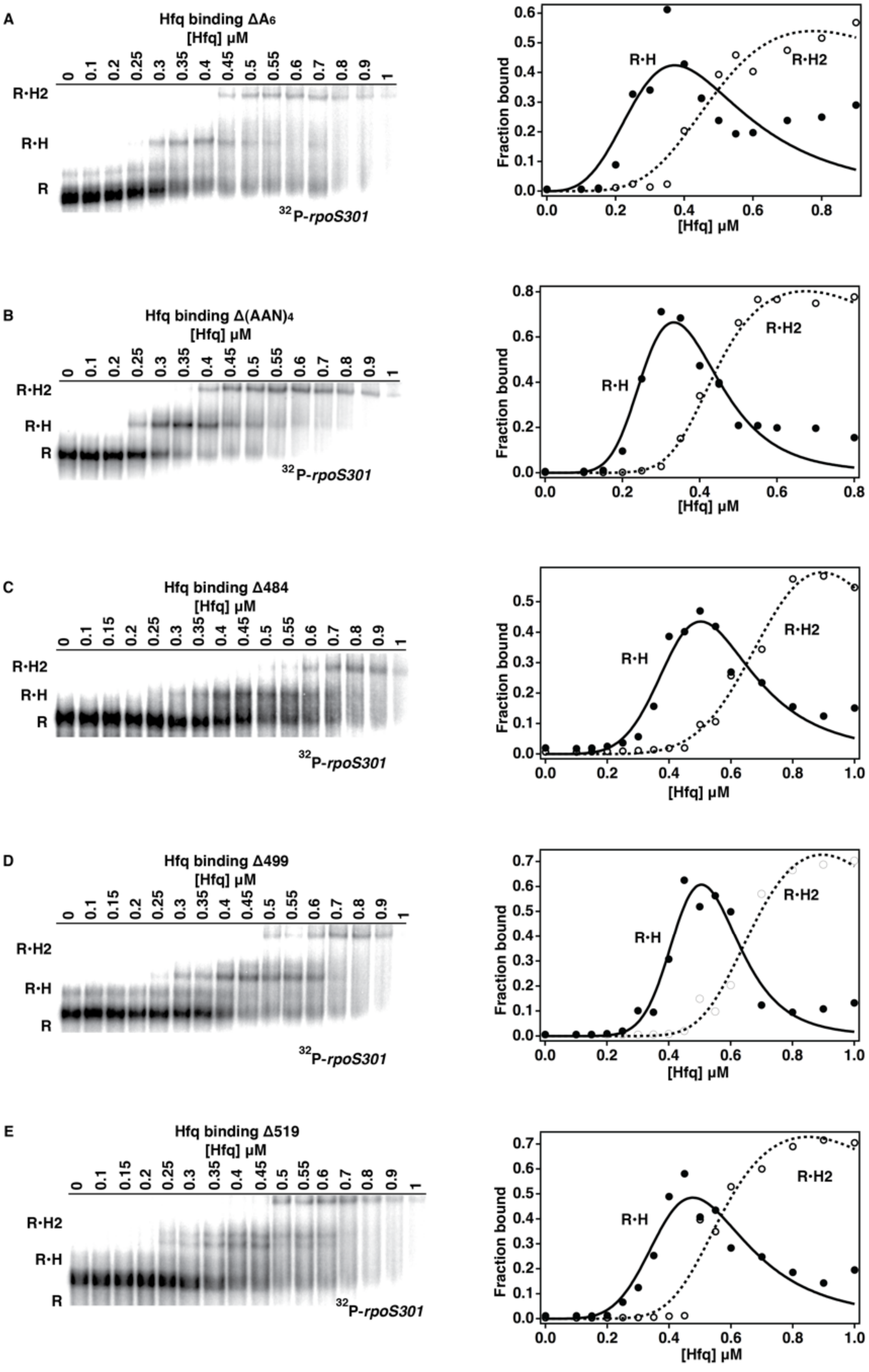


Figure 2.7. Downstream A-/U-rich motifs contribute to regulation of *rpoS* translation. (*A*) sRNA activation of *rpoS*::*lacZ* fusions *in vivo*, as in Fig. 2.2C. Hairpin loops at the positions indicated were changed to C-/G-rich sequences, in the wt *rpoS* background. Significantly different expression of mutants from the wt rpoS-lacZ fusion is marked as * ($p < 0.05$) or ** ($p < 0.01$). (*B*) Hfq binding to wt and mutated *rpoS301* RNAs was measured as in Fig. 2.3 and fit to eq. 2.2 as shown in Fig. 2.8.

Figure 2.8. Hfq binding of *rpoS301* single mutants in vitro. Binding experiments to ^{32}P-labeled rpoS301 RNA were performed as described in Fig. 2.3. (*A*) to (*E*) ΔA_6, $\Delta(AAN)_4$, Δ484, Δ499, and Δ519. Left, Hfq titration of *rpoS* resolved on native polyacrylamide gels. Right, fraction of bound *rpoS* fit to eq. 2.2. Higher molecular weight complexes were not resolved by the gel. ΔA_6 and $\Delta(AAN)_4$ bind to Hfq more tightly than Δ484, Δ499, and Δ519. Note that free ΔA_6, free Δ499, and Δ519 R•H complex migrated as two conformations. The K_d values are summarized in Table 2.1.


A
Hfq binding ΔA6
[Hfq] μM
0 0.1 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.7 0.8 0.9 1
R·H2
R·H
R
32P-rpoS301
Fraction bound
[Hfq] μM
R·H
R·H2
B
Hfq binding Δ(AAN)4
[Hfq] μM
0 0.1 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.7 0.8 0.9 1
R·H2
R·H
R
32P-rpoS301
Fraction bound
[Hfq] μM
R·H
R·H2
C
Hfq binding Δ484
[Hfq] μM
0 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.7 0.8 0.9 1
R·H2
R·H
R
32P-rpoS301
Fraction bound
[Hfq] μM
R·H
R·H2
D
Hfq binding Δ499
[Hfq] μM
0 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.7 0.8 0.9 1
R·H2
R·H
R
32P-rpoS301
Fraction bound
[Hfq] μM
R·H
R·H2
E
Hfq binding Δ519
[Hfq] μM
0 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.7 0.8 0.9 1
R·H2
R·H
R
32P-rpoS301
Fraction bound
[Hfq] μM
R·H
R·H2

Discussion

Position of the AAN motif in the *rpoS* leader affects Hfq regulation

Most bacterial mRNAs regulated by sRNAs contain AAN motifs that bind Hfq (Vogel and Luisi, 2011). The upstream AAN motifs in the *rpoS* mRNA leader are required for Hfq-dependent regulation of *rpoS* translation by sRNAs in *E. coli* (Soper et al., 2010) and accelerated sRNA-mRNA base pairing *in vitro* (Soper and Woodson, 2008; Updegrove et al., 2008). By relocating the AAN motif within the *rpoS* leader, I found that Hfq recognized this motif at all the positions tested, but Hfq-dependent regulation was only rescued when its binding site was placed close to the 5’ side of the sRNA annealing site. Therefore, the location of Hfq within the *rpoS* leader determines its chaperone function and its ability to act in translational control.

This study identifies three factors that determine how the position of the AAN motif affects Hfq regulation. First, experiments using short unstructured RNAs showed that Hfq must interact with both the sRNA and mRNA to catalyze the formation of base pairs (Panja et al., 2013) and is most effective when bound within 20 nt of the sRNA target site (Panja and Woodson, 2012). Although the *rpoS* $(AAN)_4$ motif is about 60 nt upstream of the sRNA binding site, my results suggest the flexible hinge allows the Hfq-binding domain to fold back on the inhibitory stem-loop, placing the lateral rim of Hfq where it can engage an incoming sRNA and its mRNA complement. In agreement with this hypothesis, A_{18} insertions far upstream (positions 53 and 250) or far downstream (position 519) of the sRNA binding site did not restore Hfq regulation, suggesting that the necessary tertiary interaction with the inhibitory stem was abolished.

Second, downstream A-/U-rich sequences may help position Hfq to interact with the sRNA binding site. My results show that A-/U-rich sequences at positions 484 and 499 in the *rpoS* leader contribute to Hfq regulation, as replacing them with A_{18} or mutating them reduced Hfq regulation in vivo and weakened Hfq binding in vitro. The loop at position 484 was protected by Hfq in ribonuclease footprinting experiments on a minimal *rpoS* mRNA and may interact with Hfq directly (Lease and Woodson, 2004). These nucleotides may also make RNA interactions that stabilize the overall conformation of the *rpoS* leader.

Third, A_{18} insertions in the inhibitory stem-loop of *rpoS* disturb the balance between translation repression and activation and are thus disadvantageous for sRNA regulation. This was suggested by the observation that the A_{18} insertion at the natural $(AAN)_4$ site (position 366) restored Hfq function better than the A_{18} insertion immediately 5' of the sRNA binding site (position 441), though in vitro studies predicted the latter should be more effective.

The A_{18} insertion next to the sRNA target site may hinder DsrA base pairing. In agreement with that, Δ2-441A18 barely annealed to DsrA in the absence of Hfq, although Hfq accelerated this reaction almost 60-fold. Additionally, the A_{18} insertion in the inhibitory stem-loop likely destabilizes its secondary structure, as basal expression of *rpoS* Δ2-441A18 was slightly elevated when DsrA was over-expressed in the *hfq*$^{-}$ strain (Fig. 2.2D) or in the pLac control in the *hfq*$^{+}$ strain (Fig. 2.2C). Higher basal translation initiation compensated for the reduced sRNA annealing rate, so that expression of Δ2-441A18 in the wt background was equal to expression of Δ2-366A18.

General implications for Hfq regulation of other mRNAs

In mRNAs regulated by Hfq and sRNAs, the AAN binding motif occurs at various positions with respect to the start of the open reading frame and the binding sites for sRNAs (Beisel et al., 2012; Masse and Gottesman, 2002; Prevost et al., 2007; Salim and Feig, 2010). The three factors discussed above suggest how the structural context of Hfq binding can modulate the efficiency of sRNA regulation in other mRNAs.

First, the optimal location for Hfq likely depends on whether the sRNA increases or decreases expression. In many negatively regulated mRNAs in which the sRNA represses translation by base pairing with the Shine-Dalgarno sequence, Hfq binds near the sRNA target site (De Lay et al., 2013). My results and the previous studies predict this proximity between Hfq and the sRNA target maximizes the ability of Hfq to facilitate sRNA annealing (Panja and Woodson, 2012). In some cases, Hfq binds so close to the Shine-Dalgarno sequence that Hfq directly competes with ribosome entry (Prevost et al., 2007). In studies using artificial mRNA and sRNA mimics, a stable Hfq-sRNA-mRNA ternary complex recruits RNase E to the message via interactions with a 5' monophosphate on the sRNA (Bandyra et al., 2012). Moreover, Hfq was shown to directly recruit RNase E to *ptsG* mRNA for SgrS sRNA mediated degradation (Ikeda et al., 2011). Thus, strong Hfq binding sites near the site of regulation may make translation repression or mRNA turnover more efficient.

On the other hand, positively regulated mRNAs tend to adopt more complicated secondary structures that inhibit translation initiation in the absence of sRNA but open to permit ribosome binding when the sRNA is present (Vogel and Luisi, 2011). In these examples, Hfq may bind far from the sRNA, in which case it is presumably delivered to the site of the action by the mRNA structure. The full-length *rpoS* leader is organized into distinct domains dedicated to inhibition and Hfq binding. My results suggest this organization will be generally advantageous for positive control, because Hfq can be recruited to the mRNA without disrupting its inhibitory self-structure. This organization may also allow Hfq to release the sRNA-mRNA double helix once annealing is complete.

Finally, other sequence elements in the mRNA may fine-tune or stabilize Hfq interactions. Downstream A-/U-rich patches in *rpoS* stabilize Hfq binding to the upstream $(AAN)_4$ motif and improve Hfq-dependent regulation, consistent with previous findings that Hfq restructures certain mRNAs (Geissmann and Touati, 2004; Soper et al., 2011). In *flhA* mRNA, a second Hfq binding site was identified that interacted with the proximal face of Hfq and enhanced binding of the distal face to AAN motifs (Salim and Feig, 2010). Alternatively, A-/U-rich patches in the mRNA may help displace the sRNA from the lateral rim of Hfq, where the body of the sRNA has been shown to bind (Sauer et al., 2012). Though further work is needed to fully understand how Hfq interacts with its mRNA targets, the results presented here demonstrate that the ability of Hfq to act in sRNA regulation depends on its structural context.

Materials and Methods

Construction of *rpoS* mutants and A_{18} insertions

The pUC18 plasmids of full-length wt and Δ2 *rpoS* were published previously and were used as templates for generating *rpoS* mutations and A_{18} insertions in this study (Soper et al., 2010; Soper and Woodson, 2008). The plasmid carrying wt *rpoS* was used to derive Δ484, Δ499, Δ519, wt-499A18, and wt-519A18; the plasmid carrying Δ2 *rpoS* was used to derive Δ2-366A18, Δ2-441A18, Δ2-484A18, Δ2-499A18, and Δ2-519A18. All the constructs were made by inverse PCR as described previously (Sambrook and Russell, 2006). The primers for generating mutations or insertions are listed in Appendix II.

β-Galactosidase assays

Bacterial strains (Appendix II) containing chromosomal *rpoS::lacZ* fusions were constructed from strain PM1205 as previously described (Soper et al., 2010). DNA fragments for making the recombinant bacterial strains and in vitro transcription were generated by PCR amplification of the corresponding pUC18 plasmids with primers Pbad-rpoS-F and lacZ-rpoS10aa-R (Appendix II). The *hfq*$^-$ strains were generated by P1 phage transduction with selective markers as previously described (Moore, 2011). The strains were transformed with pLac, pDsrA, pRprA, or pArcZ plasmids that were previously published (Soper et al., 2010).

Overnight cultures of transformed cells were diluted 1:500 into fresh LB medium containing ampicillin (100 μg/mL), arabinose (0.2%), and IPTG (100 μM). They were grown at 37°C with 250 rpm agitation for about 4-6 hrs or until their OD_{600} reached 0.6-0.8. The OD_{600} was measured, and the cultures were lysed and assayed for β-galactosidase activity as described previously (Majdalani et al., 1998). The reaction velocities were determined using a microplate reader (Molecular Device Thermomax), and the specific activity was calculated as V_{max}/OD_{600}. The V_{max} of each culture was the average of 3 aliquots from the same culture, and the reported specific activity of the strain was the average of at least 3 independent experiments.

RNA Preparation

In vitro transcription templates of *rpoS301* were generated by PCR amplification of the corresponding pUC18 plasmids with primers rpoS301F and rpoS576R (Appendix II). The unlabeled RNAs were transcribed with T7 RNA polymerase from the PCR templates described above. The RNAs were purified by denaturing PAGE and recovered by phenol-chloroform extraction and ethanol precipitation, as described previously (Zaug et al., 1988). The ^{32}P labeled *rpoS301* was transcribed in vitro in the presence of $\alpha^{32}P$-ATP (Lease and Woodson, 2004; Soper and Woodson, 2008) and purified through Chroma spin+ TE-100 columns (Clontech Laboratories).

Hfq Preparation

Wild-type *E.coli* Hfq was overexpressed and harvested as previously described (Zhang et al., 2002). The cells were homogenized (Avestin EmulsiFlex-C3) in 50 mL lysis buffer (50 mM HEPES pH 7.5, 0.5 M NaCl, 20 mM imidazole, 5% glycerol), followed by 1 hr DNase I (50 µl of 2000 units/µl) treatment on ice. The cell lysate was centrifuged and filtered prior to loading on a prepared 5 mL Hi-Trap Co^{2+} column. The column was washed with 50 mL lysis buffer and 150 mL wash buffer (50 mM HEPES pH 7.5, 1 M NaCl, 20 mM imidazole, 5% glycerol), then eluted with 25 mL wash buffer containing 250 mM imidazole. The eluate was concentrated to 5 mL and dialyzed overnight against 1 L 20 mM Na-HEPES pH 7.5, 100 mM NaCl, 0.5 mM EDTA. The protein sample was loaded on a UNO S6 ion exchange column (Bio-Rad), washed with dialysis buffer at 2 mL/min flow rate for 30 min and eluted with a linear gradient of 0.1 M to 1 M NaCl. The desired fractions were pooled and dialyzed against 1 L of Hfq storage buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 250 mM NH_4Cl, 10% glycerol).

Hfq Binding

The equilibrium binding reactions with Hfq and rpoS301 RNA were assembled as previously described (Soper and Woodson, 2008). After 10 min at room temperature, 2 μL aliquots were loaded on a native 6% polyacrylamide gel in 1X THEM2 (66 mM HEPES, 34 mM Tris, 0.1 mM EDTA, 2 mM $MgCl_2$). The fractions of bound *rpoS*, f_{RH}, were quantified as previously described (Lease and Woodson, 2004) and were fit with a partition function (IGOR Pro, WaveMetrics). For *rpoS* constructs that formed only one specific complex with Hfq, I assumed two unequal independent binding sites

$$f_{RH}=\frac{([Hfq]/K_1)^n}{Q_{RH}} \tag{2.1a}$$

$$Q_{RH}=1+([Hfq]/K_1)^n+([Hfq]^2/K_1K_2)^n \tag{2.1b}$$

in which [Hfq] is the concentration of Hfq monomers, K_1 is the dissociation constant of the specific complex, K_2 is the apparent dissociation constant of nonspecific binding, and n is the Hill coefficient. Similarly, for *rpoS* constructs that formed two specific Hfq complexes, R•H and R•H2 fractions were fit to the second and third terms of a partition function assuming three unequal independent binding sites:

$$f_{RH}=\frac{([Hfq]/K_1)^n}{Q_{RH}} \tag{2.2a}$$

$$f_{RH2}=\frac{([Hfq]^2/K_1K_2)^n}{Q_{RH}} \tag{2.2b}$$

$$Q_{RH}=1+([Hfq]/K_1)^n+([Hfq]^2/K_1K_2)^n+([Hfq]^3/K_1K_2K_3)^n \tag{2.2C}$$

DsrA annealing

Reactions to measure the annealing kinetics of DsrA and *rpoS301* were assembled as previously described (Soper and Woodson, 2008), except that the native gel was done in 1X THEM2 buffer as described above. The fractions of bound *rpoS* were quantified as previously described (Lease and Woodson, 2004) and were fit with either a single or double exponential rate equation (IGOR Pro, WaveMetrics)

$$f_{RD} = f(1\text{-}\exp(-k_{obs1}t)) \tag{2.3}$$

$$f_{RD+RDH} = f_1(1\text{-}\exp(-k_{obs1}t)) + f_2(1\text{-}\exp(-k_{obs2}t)) \tag{2.4}$$

in which f , f_1, and f_2 are the fractions of the reaction that followed the corresponding annealing rate, t is the annealing time, k_{obs1} is the apparent annealing rate for the slow phase reaction, k_{obs2} is the apparent annealing rate for the fast phase reaction.

Chapter 3 Hfq regulation depends on binding to the A-rich and U-rich sequences in the *rpoS* and *flhD* leaders

Introduction

Though Hfq is well known to bind specific sequences in sRNAs and mRNAs, how it restructures its targets for translational control is not known. Soper et al. showed that an (AAN)$_4$ motif upstream of the sRNA target site binds the distal face of Hfq and recruits Hfq to the mRNA for sRNA annealing and regulation (Soper et al., 2010; Soper and Woodson, 2008). Nevertheless, Lease et al. and I both used mutagenesis experiments to show that Hfq could also interact with a 5' UUAUUU "U$_5$" loop downstream of the sRNA (Lease and Woodson, 2004; Peng et al., 2014), raising the possibility that Hfq contacts more than one region of the *rpoS* leader. Furthermore, Soper et al. previously found that Hfq restructures the *rpoS* leader, making the sRNA binding site more accessible to anti-sense probes (Soper et al., 2011). It is plausible Hfq alters RNA structure by binding to multiples sites, as observed for RprA and OxyS sRNAs (Henderson et al., 2013).

To test how Hfq interacts with different mRNA leaders, I compared Hfq binding and translational regulation of *rpoS* and *flhD* leaders. *rpoS* encodes s^S, a major stress-response regulatory in *E. coli* that is up-regulated by at least three different sRNAs (Hengge-Aronis, 2002). *flhDC* mRNA encodes the master regulator of flagella synthesis, which is positively and negatively regulated by direct pairing to five sRNAs (De Lay and Gottesman, 2012).

SHAPE footprinting confirmed that the Hfq distal face binds to AAN motifs in both mRNA leaders. Moreover, the Hfq lateral surface also interacts with a U-rich motif in the *rpoS* leader to position and partially unwind an inhibitory stem that contains the sRNA binding site and that blocks ribosome binding. By contrast, this U-rich interaction is absent in the *flhD* leader that shows minimal structural change upon Hfq binding. *In vitro* sRNA annealing and *in vivo* translation regulation of *rpoS* and *flhD* leaders reveals the importance of the U-rich motif for effective Hfq activity.

Results

Hfq binds A-rich and U-rich motifs in *rpoS* mRNA

I used SHAPE footprinting to identify Hfq interaction sites in the *rpoS* leader RNA. For these experiments, I used *rpoS301*, a 284-nt variant of the 576-nt *rpoS* leader that lacks a non-essential upstream domain but retains the Hfq binding domain and inhibitory stem-loop needed for translational control and Hfq and sRNA binding (Peng et al., 2014). The *rpoS301* RNA folds homogeneously *in vitro* and has a similar SHAPE modification pattern as the full-length *rpoS* leader (Fig. 3.2A and B), indicating that *rpoS301* retains the native secondary structure.

I first compared the SHAPE modification levels of free *rpoS301* RNA with the RNA bound to DsrA sRNA or Hfq (Fig. 3.1A and Fig. 3.2C). I then categorized the decrease or increase in relative SHAPE reactivity based on a histogram of the entire data set (Fig. 3.2D), which reflects either the accessibility of the ribose 2’OH or the flexibility of the RNA backbone (McGinnis et al., 2012). As expected, base pairing between *rpoS* mRNA and DsrA sRNA protected the DsrA binding site in the inhibitory stem from modification, reducing the SHAPE reactivity by ~30-40% (Fig. 3.1B). However, the SHAPE reactivity of the upstream and downstream domains did not change appreciably, suggesting they are unaffected by DsrA (Fig. 3.1A, green trace and Fig. 3.1B).

By contrast, Hfq binding remodeled the *rpoS* mRNA structure extensively (Fig. 3.1A, magenta trace and Fig. 3.1C). First, Hfq increased the reactivity of the inhibitory stem and the helix connecting the $(AAN)_4$ and A_6 motifs 2-3 fold compared to the free RNA. These residues were uniformly and moderately modified in the *rpoS*•Hfq complex, suggesting that Hfq partially opens the mRNA secondary structure. This Hfq-induced structural change in the inhibitory stem was also reported based on RNase footprinting experiments (Hammerle et al., 2013).

Second, Hfq binding caused distinct hyper-reactive nucleotides in three regions: the $(AAN)_4$ motif in the Hfq binding domain, the U_5 motif in the downstream domain, and A157 in the inhibitory stem. The first A of every AAN triplet was 4-9 times more modified in the Hfq complex than in the RNA control (Fig. 3.1A, nt. A80, A83, and A85; Fig. 3.2C, top left). Because the $(AAN)_4$ motif is known to bind the distal face of Hfq (Soper et al., 2010; Soper and Woodson, 2008). I suspected that Hfq locks the RNA backbone in a highly reactive conformation. Indeed, this enhanced reactivity was explained by the crystal structure of *E. coli* Hfq in complex with rA_{18}, in which the first A adopts a 2’-endo ribose conformation, positioning the 2’OH close to the P-O vector (Fig. 3.2E) (Link et al., 2009).

Third, the third and fourth residues in the U_5 motif were 8 times more modified in the *rpoS*•Hfq complex (Fig. 3.1A magenta trace residue 194A and 195U; Fig. 3.2C bottom left panel), indicating that Hfq also interacts with this U-rich loop as suggested by previous RNase footprinting experiments (Lease and Woodson, 2004). Lastly, A157 5' of the sRNA binding site in the inhibitory stem was 6 times more reactive in the *rpoS*•Hfq complex (Fig. 3.1A magenta trace), revealed a previously undiscovered interaction between Hfq and the start of the inhibitory domain.

Importantly, the modification pattern of the DsrA•*rpoS*• Hfq ternary complex showed that Hfq releases the inhibitory stem and U_5 motif but remains bound to the upstream $(AAN)_4$ motif (Fig. 3.1D). This is consistent with *in vitro* annealing experiments showing that Hfq cycles off the sRNA-mRNA duplex after the RNAs have base paired (Lease and Woodson, 2004). DsrA protected its target site from modification as expected (Fig. 3.1A, blue trace and Fig. 3.1D). Moreover, this region was ~50-80% less modified in the ternary complex than in the DsrA•*rpoS* complex, consistent with tighter DsrA•*rpoS* binding in the presence of Hfq (Soper et al., 2010). Meanwhile, nucleotides upstream of the Shine-Dalgarno sequence became 2-3 fold more reactive, or accessible, in the ternary complex.

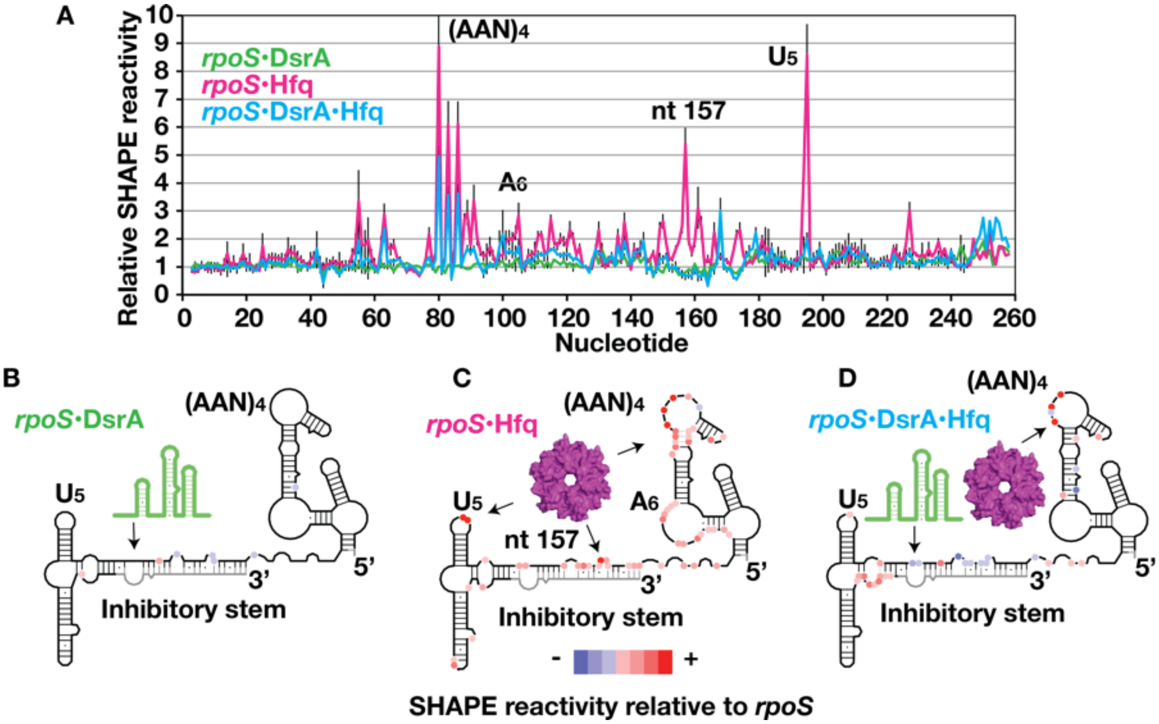
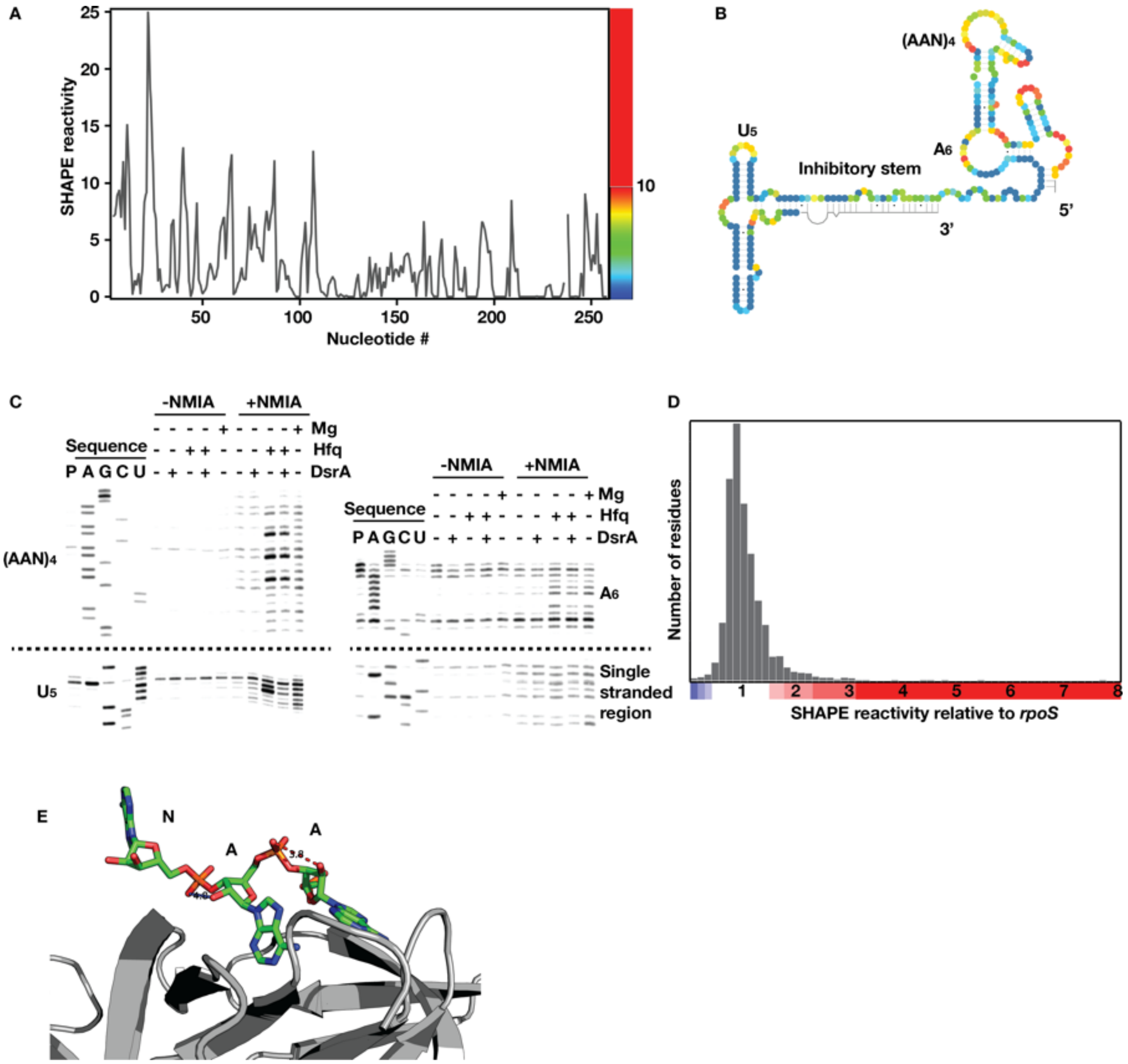


Figure 3.1. Conformational changes in *rpoS* mRNA from binding of DsrA and Hfq. (*A*) SHAPE reactivity of 50 nM *rpoS301* RNA in complex with 200 nM DsrA (green trace), 333 nM Hfq (magenta), or DsrA and Hfq (blue) relative to *rpoS* RNA alone (Fig. 3.2). $(AAN)_4$ motif, nt. 77-88; inhibitory stem, nt. 149-184 and 249-284; U_5 motif, nt. 192-197. NMIA modification was carried out at 37 ˚C (see Methods) and the extent of modification was measured by primer extension (Fig. 3.2C). Error bars represent ± S.D. for at least three independent experiments. (*B-D*) Schematic of SHAPE reactivity relative to free *rpoS* RNA for each complex, from a histogram of the entire data set (Fig. 3.2D). Red circles, nucleotides with enhanced SHAPE reactivity; blue circles, nucleotides with reduced SHAPE reactivity; black line, regions with unchanged SHAPE reactivity; grey line, regions with no SHAPE data. Arrows indicate DsrA and Hfq binding regions in *rpoS* RNA.

Figure 3.2. SHAPE footprinting of *rpoS* mRNA with sRNA and Hfq. (*A*) Modification of ribose 2′OH was quantified by primer extension and analyzed by sequencing gel electrophoresis. To determine the overall secondary structure, traces from primer extensions covering different regions of the *rpoS301* RNA were overlapped. (*B*) *rpoS301* secondary structure consistent with experimental SHAPE modification data (colors). The linker between the upstream and downstream domains is predicted MC-Sym to fold into a short stem-loop, in agreement with its low reactivity. Gray, no data. (*C*) Selected sequencing gel images showing hyper-reactive residues induced by Hfq binding at $(AAN)_4$ motif (left top panel) and U_5 motif (left bottom panel). Hfq also opens the secondary structure at the A_6 motif (right top panel) but does not affect other single-stranded regions (right bottom panel). (*D*) Histogram of SHAPE reactivity relative to *rpoS* RNA for the entire data set. Values below 0.6 (~3% of nt) were considered protected; values above 1.75 were considered enhanced (~10% of nt). Nucleotides were clustered based on their relative SHAPE reactivity: dark to light blue, protected, 0.140-0.287 (~0.2%), 0.287-0.434 (~0.3%), and 0.434-0.581 (~0.6%); white, unchanged, 0.728-1.757 (~87.1%); light to dark red, enhanced, 1.757-2.051 (~4.5%), 2.051-2.639 (~3.4%), 2.639-3.521 (~1.5%), and 3.521-8.813 (~0.8%). RNA secondary structure schematics were drawn with XRNA (http://rna.ucsc.edu/rnacenter/xrna/xrna.html). (*E*) Structure of AAN triple bound to the distal face of Hfq (pdb 3GIB) showing the short 3.8 Å distance between the first 2′OH and the phosphate of the second A. This conformation is associated with strong reactivity toward electrophiles used for SHAPE chemistry (Steen et al., 2012).


A
SHAPE reactivity
25
20
15
10
5
0
50
100
150
200
250
Nucleotide #
10
B
(AAN)4
U5
A6
Inhibitory stem
3'
5'
C
-NMIA
+NMIA
Mg
Hfq
DsrA
Sequence
P A G C U
(AAN)4
U5
A6
Single stranded region
D
Number of residues
1
2
3
4
5
6
7
8
SHAPE reactivity relative to rpoS
E
N
A
A
4.0
3.8

A U_5 motif binds the lateral rim of Hfq

To test which surfaces of Hfq contact *rpoS* mRNA, I repeated the SHAPE experiments with Hfq mutants Y25D, R16A, and K56A that disrupt RNA binding to the distal face, lateral rim, and proximal face, respectively (Mikulecky et al., 2004; Sauer et al., 2012; Updegrove and Wartell, 2011). As expected, the Y25D mutation specifically disturbed the hyper-reactive modification of the $(AAN)_4$ motif (Fig. 3.3A), consistent with binding to Hfq's distal face (Link et al., 2009; Soper et al., 2011). Nonetheless, this mutation only partially impairs binding, as it eliminated the hyperactivity of only the third (AAN) triplet, and shifted the modification pattern one nucleotide upstream (Fig. 3.3A, purple trace for residues C79 and C82). The SHAPE reactivity of the inhibitory stem and the downstream domain were the same as in the WT Hfq complex (purple and magenta traces), indicating that those regions interact with a different surface of Hfq.

Strikingly, the R16A rim mutation abolished interactions with the inhibitory stem and the U_5 motif while leaving interactions with the $(AAN)_4$ motif intact (orange trace, Fig. 3.3B). The lost hyperactivity of the U_5 motif (A194 and U195) suggested that this motif directly contacts the lateral rim of Hfq. Modification of A157 returned to the average level, and modification of C137, C140, and C165 increased ~3-fold (orange and magenta traces), indicating that the perturbed interaction with the rim also changed the conformation of the inhibitory stem. Finally, K56A mutant did not appreciably change the modification pattern (green trace, Fig. 3.3C), confirming that the proximal face of Hfq does not bind *rpoS* mRNA directly.

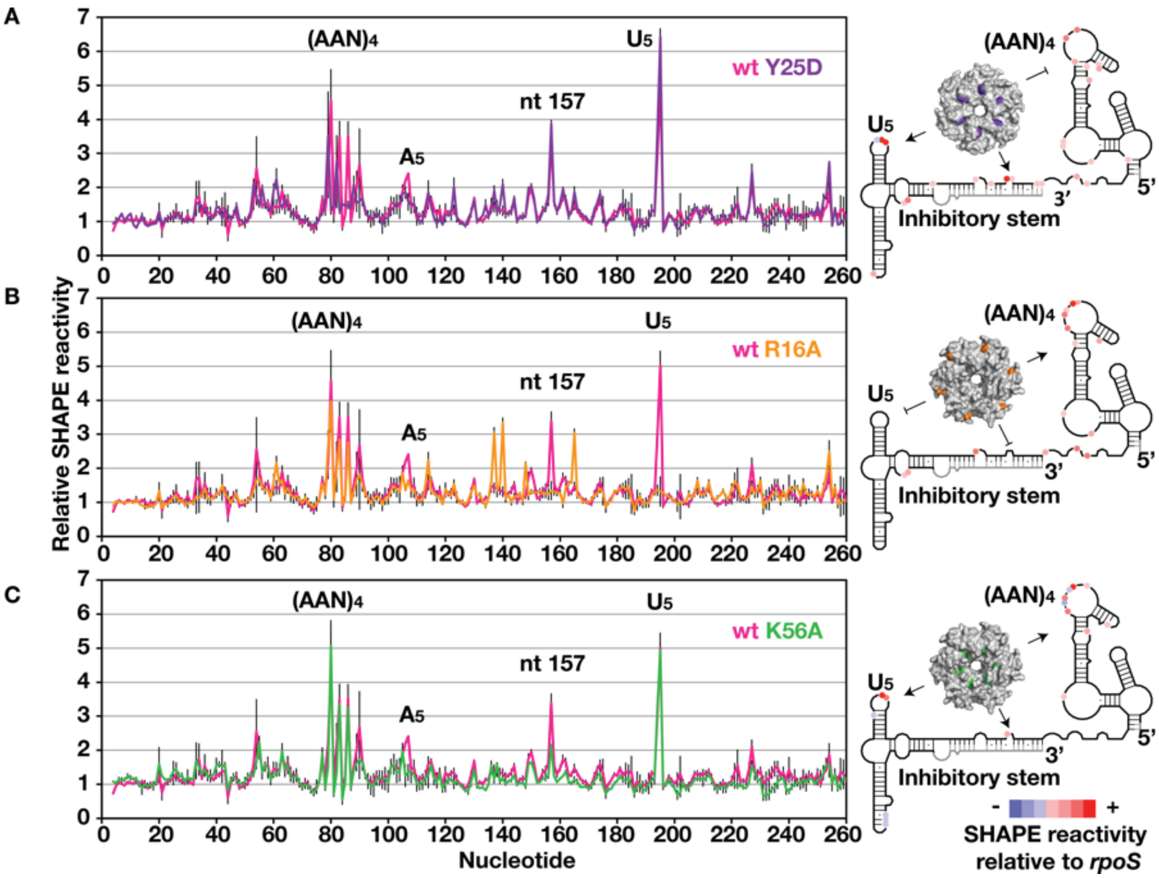


Figure 3.3. Mapping *rpoS* interaction sites on Hfq. Left, SHAPE modification of *rpoS* RNA in complex with wt Hfq (magenta trace) or an RNA binding surface mutation. Error bars as in Fig. 3.1. (*A*) Distal face Y25D mutation (purple) disrupts Hfq binding to (AAN)$_4$ motif. (*B*) Rim R16A mutation (orange) disrupts Hfq interactions with the inhibitory stem and U$_5$ motif. (*C*) Proximal face K56A mutation (green) retains the hyper-modification of the three direct binding sites, but did not cause an appreciable change in the RNA secondary structure.

U_5 motif binding at Hfq lateral surface facilitates DsrA annealing

The SHAPE results showed that the lateral surface of Hfq directly contacts the downstream U_5 motif in the *rpoS* mRNA, while the distal face remains bound to the upstream $(AAN)_4$ motif. To investigate whether the U_5 motif is required for regulation of *rpoS* translation by Hfq and sRNAs, I tested whether mutations in the U_5 stem-loop alter expression of *rpoS::lacZ* fusions in *E. coli* and DsrA annealing *in vitro* (Fig. 3.4). I created three mutations in the U_5 stem-loop, either replacing the UUAUUU loop with UCGC (Fig. 3.4A, ΔU_5), shortening the stem by 3 bp (Fig. 3.4A, U_5SS), or expanding the loop by an extra 9 nt (Fig. 3.4A, U_5UL). All three mutations diminished the ability of DsrA and RprA sRNAs to up-regulate expression of full-length *rpoS*::lacZ fusions in the *E. coli* chromosome by 20-40% (Fig. 3.4B, green and gold bars). The magnitude of this effect was similar to that of mutating the upstream $(AAN)_4$ and A_6 motifs (Fig. 3.5B, Δ2). The U_5 mutations had a smaller effect on up-regulation by ArcZ sRNA (Fig. 3.4B, red bars).

To investigate whether the U_5 motif is important for DsrA annealing in vitro, I next measured the stability of the DsrA--*rpoS* complex using gel mobility shift assays (Peng et al., 2014). I titrated ^{32}P-labeled *rpoS* mRNA with DsrA sRNA (0 - 2 μM) (Fig. 3.5C) and quantified the total fraction of *rpoS*•DsrA and *rpoS*•DsrA•Hfq complexes as a function of DsrA concentration (Fig. 3.4C). Without Hfq, the ΔU_5 mutation did not change the strength of the DsrA-*rpoS* RNA interaction, suggesting that this mutation does not alter the structure of free *rpoS* mRNA (Table 3.1). With Hfq present, however, the ΔU_5 mutation stabilized DsrA binding twofold (Table 3.1) compared to the WT *rpoS*•DsrA•Hfq ternary complex. The ΔU_5 mutation also increased dissociation of Hfq from the base paired *rpoS*•DsrA RNAs (Fig. 3.4D).

I next measured Hfq's ability to increase the rate of DsrA annealing with *rpoS* mutant mRNAs (Fig. 3.4D and 3.6D). As expected, in the absence of Hfq, both WT and ΔU_5 *rpoS* mRNA base paired with DsrA at similar rates (0.026 min^{-1} and 0.031 min^{-1}, respectively; Table 3.1). In the presence of Hfq, a lower proportion of ΔU_5 than WT *rpoS* mRNA annealed with DsrA during the first 30 s (Table 3.1). Because only ternary DsrA•*rpoS*• Hfq complex is formed at the beginning of the reaction (Fig. 3.4D top), this difference reflects the ease of binding DsrA, and not the rate of Hfq release from the *rpoS*•DsrA duplex. Thus, these results suggest that interactions between Hfq and the U_5 motif distort the *rpoS* mRNA conformation for efficient DsrA entry.

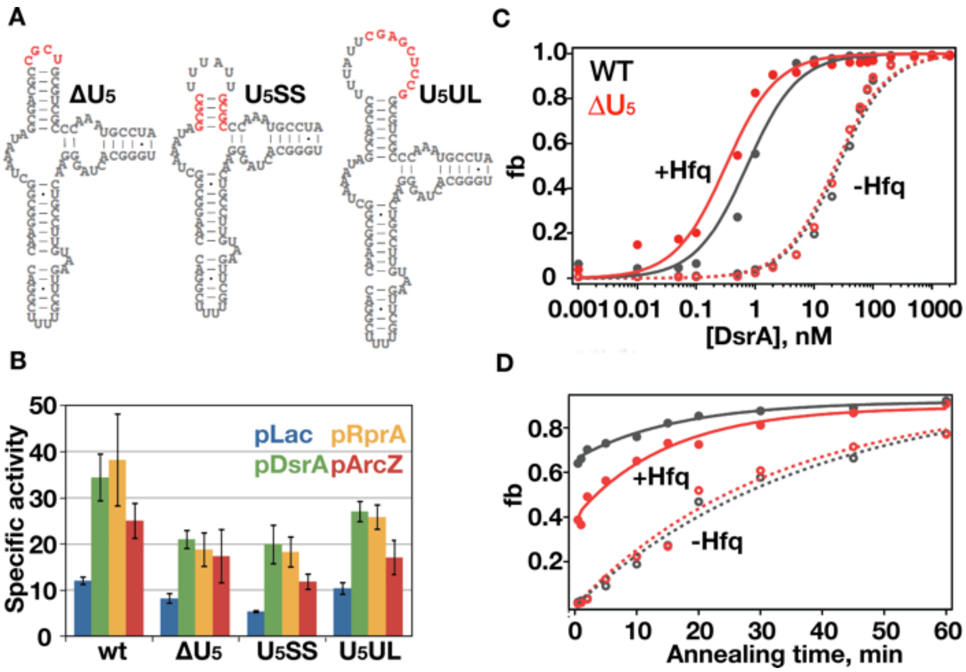


Figure 3.4. Function of U_5 motif in sRNA binding and regulation. (*A*) Mutations in the U_5 motif (red) delete the U-rich sequence in the loop (ΔU_5), shorten the helix (U_5SS), or insert a GC-rich sequence in the loop (U_5UL). (*B*) β-galactosidase activity assays measure translation of *rpoS::lacZ* in *E. coli* when sRNAs are overexpressed from IPTG-inducible plasmids. Empty pLac vector (blue); pDsrA (green); pRprA (orange); pArcZ (red). (*C-D*) DsrA binding to WT *rpoS* RNA (gray) and ΔU_5 *rpoS* RNA (red) without Hfq (open circles and dashed lines) or with Hfq (solid circles and solid lines). (*C*) Equilibrium binding between *rpoS* and DsrA RNAs, measured by native gel mobility shift. Fraction of *rpoS*•DsrA (RD) or *rpoS*•DsrA•Hfq (RDH) versus [DsrA] was fit to a single-site binding isotherm. (*D*) *rpoS*-DsrA annealing kinetics. Data were fit to single (no Hfq) or double (+Hfq) rate equations (Lease and Woodson, 2004). Error bars represent ±S.D. for at least three independent experiments.

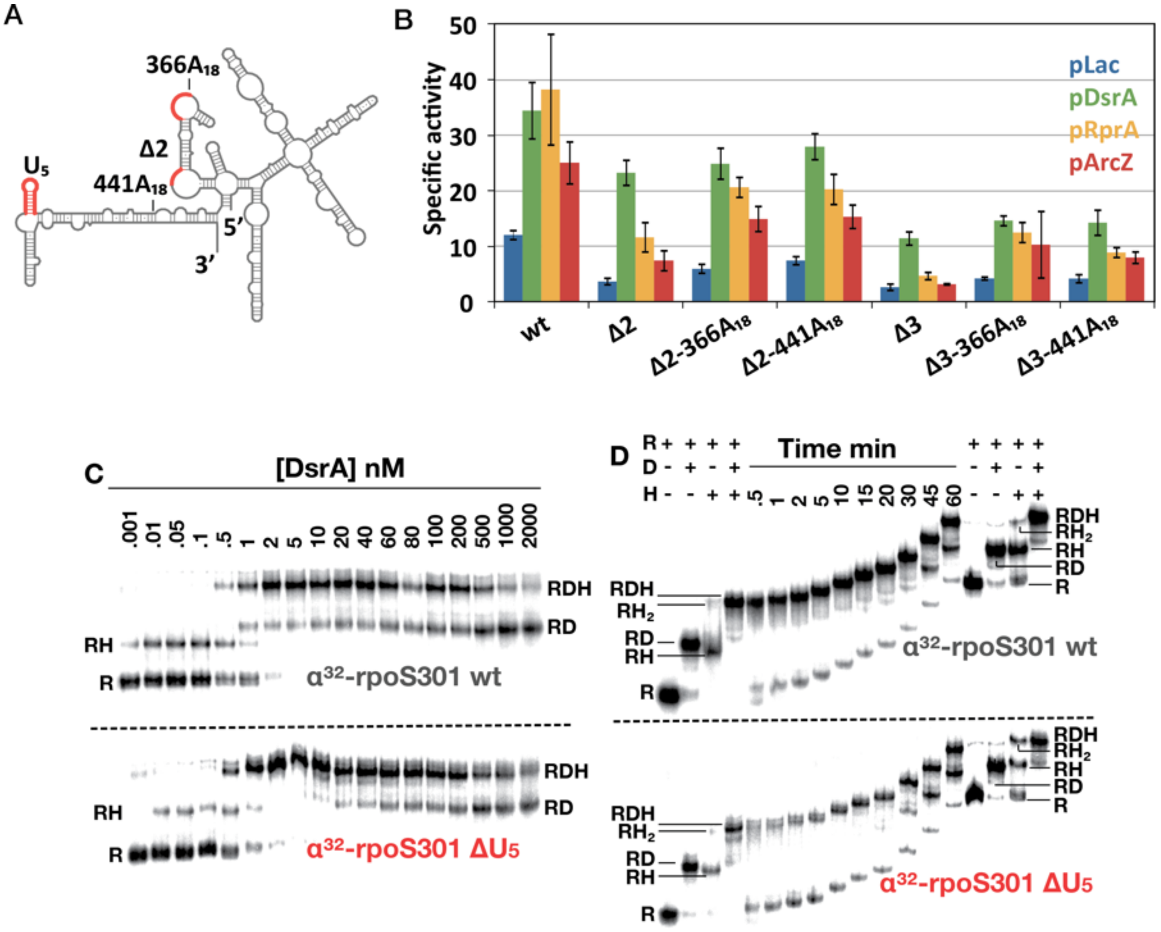


Figure 3.5. The U5 motif functions independently of AAN motif. (*A*, *B*) Full-length *rpoS* leader (576 nt) fused to *lacZ* was used to measure in vivo expression of β-galactosidase as in Fig. 3.4B. Δ2, upstream (AAN)$_4$ and A$_6$ motifs were replaced with a GC-rich sequence (Soper et al., 2010); Δ3, ΔU$_5$ plus Δ2 mutation; 366A$_{18}$ and 441A$_{18}$ contain an A18 insertion at position 366 or 441 to rescue Δ2 or Δ3 *rpoS* function (Peng et al., 2014). (*C*, *D*) Native polyacrylamide gel mobility shift assay for DsrA binding to *rpoS* mRNA. (*C*) Equilibrium binding at different DsrA concentrations. Fraction bound was calculated from counts in each lane in *rpoS*•DsrA (RD) or *rpoS*•DsrA•Hfq (RDH) bands. (*D*) Binding kinetics, as in (*C*). Samples were loaded at various times during electrophoresis.

Table 3.1 Summary of *rpoS*•DsrA binding constant and annealing rate

	DsrA binding constant (nM^{-1})		DsrA annealing rate (min^{-1})	
	-Hfq	+Hfq	-Hfq	+Hfq
wt	25.69 ± 2.24	0.72 ± 0.10	0.03 ± 0.005 (98.1%)	6.37 ± 1.27 (fast phase 65.8%) 0.06 ± 0.01 (slow phase 26.0%)
ΔU_5	22.68 ± 1.82	0.32 ± 0.04	0.03 ± 0.006 (93.6%)	4.57 ± 2.42 (fast phase 39.8%) 0.07 ± 0.02 (slow phase 49.6%)

Summary of *rpoS*•DsrA binding constants and *rpoS*•DsrA annealing rate measured by native polyacrylamide gel electrophoresis. The values was generated from data shown in Fig. 3.4C and D. The equations and calculations were described previously (Peng, Soper & Woodson 2014).

Further SHAPE footprinting results confirmed that this defect in DsrA annealing was due to impaired Hfq binding at the U_5 motif, based on the loss of hyper-reactivity at this position (Fig. 3.6A). The ΔU_5 mutation also reduced the SHAPE reactivity of the upstream $(AAN)_4$ motif by ~80% compared to the WT mRNA (Fig. 3.6A), consistent with a reduction in Hfq affinity (Peng et al., 2014). Surprisingly, I still observed strong modification of A157 in the inhibitory stem, suggesting this contact depends on recruitment of Hfq to the $(AAN)_4$ motif rather than U_5 motif. By contrast, the SHAPE reactivity of the inhibitory stem was no longer enhanced, consistent with the previous conclusion that the U_5 motif is needed for Hfq to open the inhibitory stem. The U_5SS and U_5UL mutations also disrupted Hfq binding at the U_5 motif (Fig. 3.6B and C), in agreement with the lower *in vivo* expression of *rpoS::lacZ* observed in these mutations.

Finally, I compared the expression of *rpoS::lacZ* fusions lacking the upstream $(AAN)_4$ and A_6 motifs (Δ2) to a variant lacking all three Hfq binding sites (Δ3; Fig. 3.5A and B). Loss of the U_5 motif further reduced up-regulation of *rpoS* expression by 50%, showing that the $(AAN)_4$ and U_5 motifs not only interact with different surfaces of Hfq, but have distinct functions in the regulation of *rpoS* translation by sRNAs and Hfq.

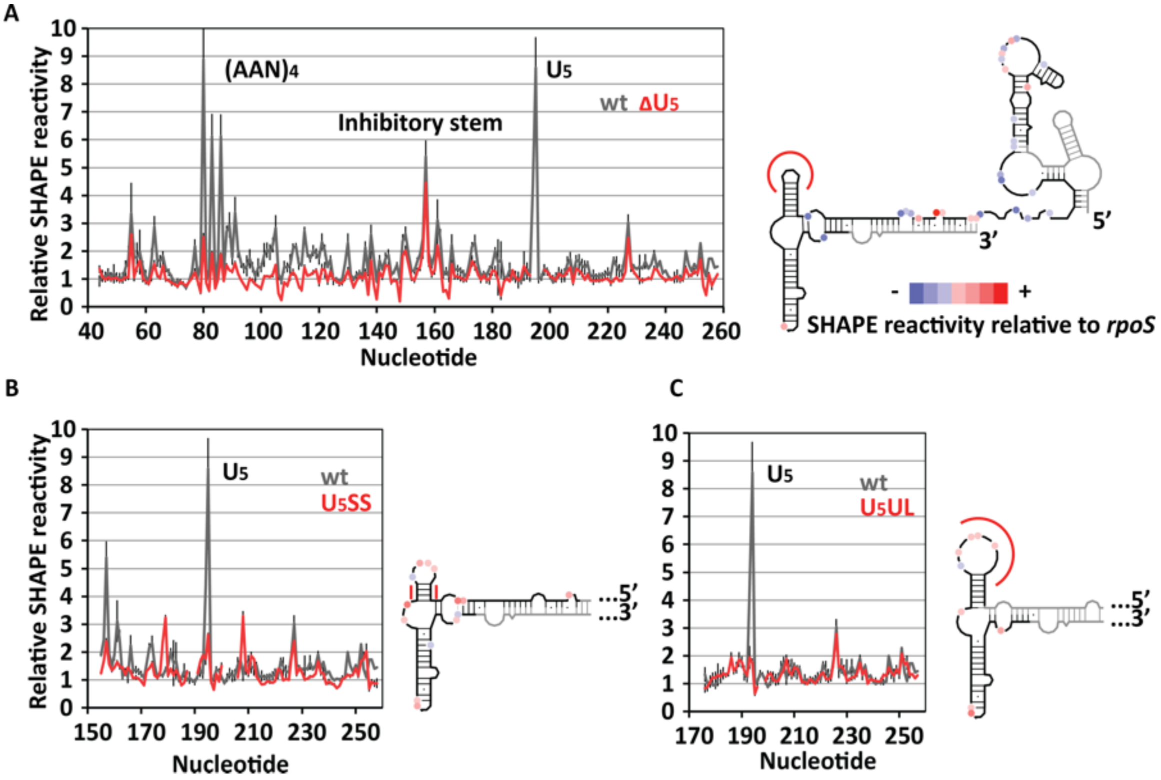


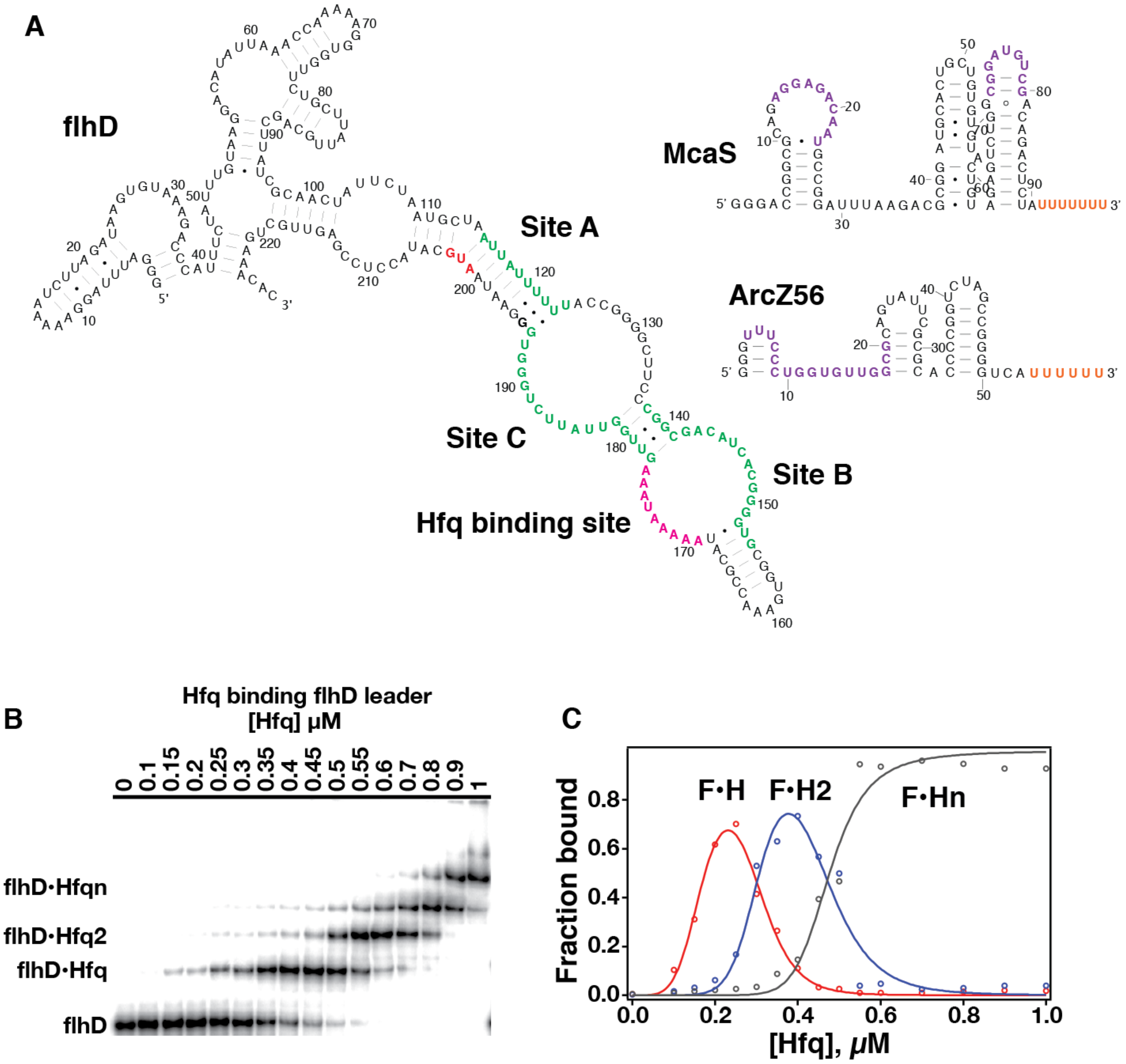
Figure 3.6. Hfq specifically recognizes the U_5 motif. Comparison of relative SHAPE reactivity of WT *rpoS*•Hfq complex (grey trace) and U_5 mutant *rpoS*•Hfq complexes (red trace). Sites of mutations are highlighted by red lines on the schematics, as shown in Fig. 3.4A.

Hfq only binds the A-rich motif and does not remodel *flhD* structure

To investigate if a U-rich motif is also required for Hfq to restructure other mRNA leaders, I used SHAPE footprinting to identify Hfq binding sites in the *flhD* leader. The full-length *flhD* leader (nt 1- 225, *flhD27*) that contains the entire 5’-UTR (198 nt) and the first 9 codons (27 nt) is sufficient for sRNA regulation of *flhDC* translation *in vivo* (De Lay and Gottesman, 2012). This RNA fragment also formed a homogeneous conformation *in vitro* as compared to the shorter truncations, including *flhD6*, *flhD9*, and *flhD14* (sequence listed in Appendix II), because *flhD27* RNA migrated as a single band during the native polyacrylamide gel electrophoresis (Fig. 3.7B and data not shown).

Like the *rpoS* leader, the *flhD* leader RNA bound multiple Hfq hexamers (Fig. 3.7B). The fraction bound was fit with a partition function assuming a tight, a weak, and nonspecific binding sites (Chapter 2 Material and Methods eq. 2.2). The specific Hfq binding constants of the *flhD* leader are comparable to those of the *rpoS* leader (a tight binding constant of ~ 0.2 μM and a weak binding constant of ~ 0.4 μM) (Fig. 3.7 C).

Figure 3.7. *flhD* mRNA interacts with Hfq and McaS and ArcZ56 sRNAs. (*A*) The full-length *flhD* leader (227 nt) was predicted to form an inhibitory secondary structure based on minimum free energy (SHAPE data not included) by the Mfold webserver (Zuker, 2003). The RNA contains an AAN motif that can bind to the Hfq distal face (magenta) and sRNA annealing sites A-C (green). McaS and ArcZ56 sRNAs contains *flhD* annealing sites (colored in purple) and a 3'-U tail that can interact with the Hfq proximal face (colored in orange). (*B*) Hfq titrations of uniformly labeled *flhD* leader at 25ºC. Complexes of *flhD* mRNA (flhD) with one or more Hfq multimers are labeled flhD•Hfq, flhD•Hfq2, and flhD•Hfqn. (*C*) Fraction of bound *flhD* leader as a function of [Hfq] was fit to a partition function (Materials and Methods). Red, F•H (k_1); blue, F•H2 (k_2); and grey, F•Hn (k_n). The fraction of F•H and F•H2 decreased at high [Hfq] due to the formation of high molecular weight (non-specific) complexes.

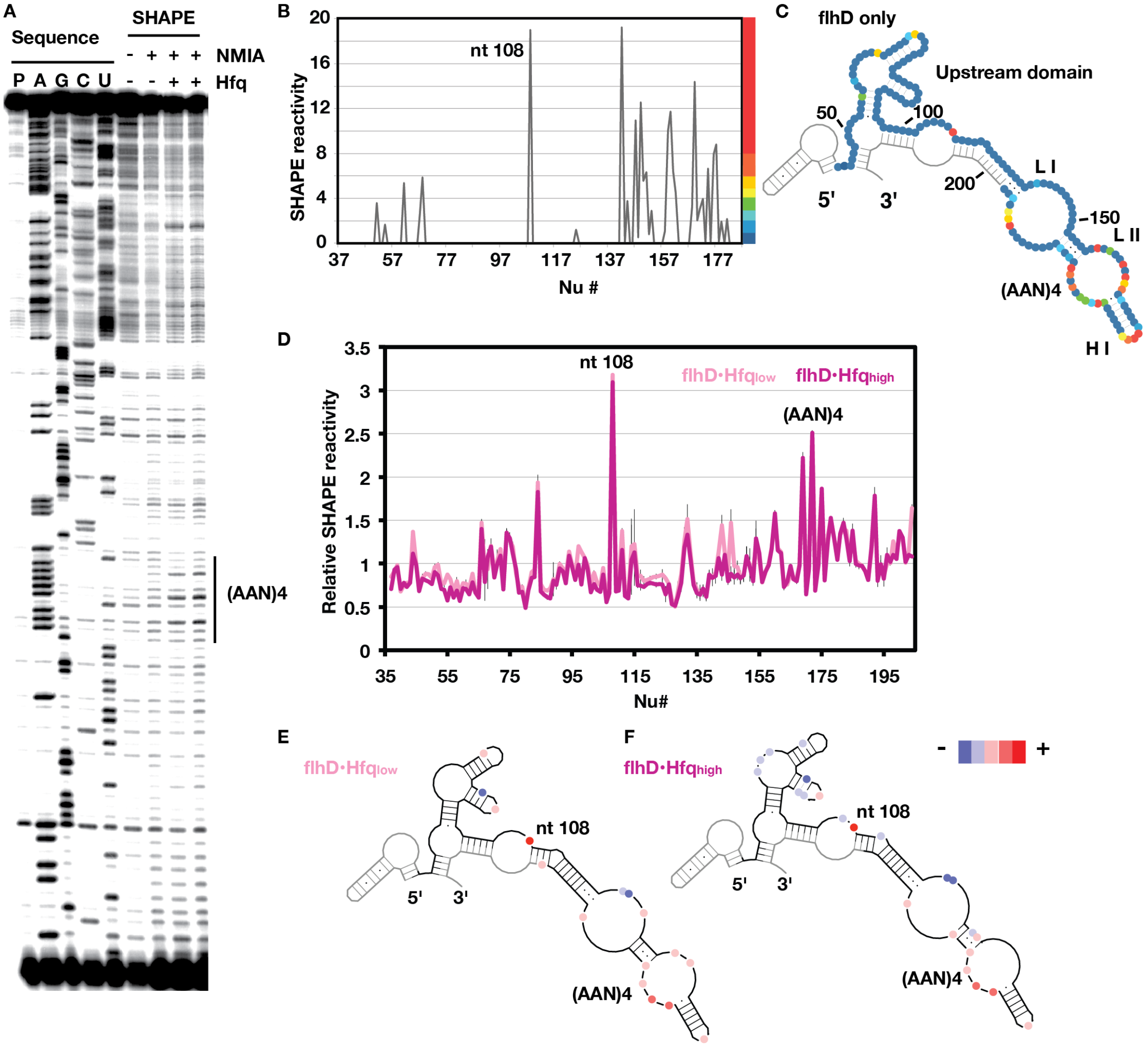

A
flhD
McaS
ArcZ56
Site A
Site B
Site C
Hfq binding site
5'
3'
B
Hfq binding flhD leader
[Hfq] μM
0
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.7
0.8
0.9
1
flhD•Hfqn
flhD•Hfq2
flhD•Hfq
flhD
C
F•H
F•H2
F•Hn
Fraction bound
0.0
0.2
0.4
0.6
0.8
1.0
[Hfq], μM

I then measured SHAPE modification of the *flhD* leader in the absence and presence of Hfq (Fig. 3.8). The secondary structure of *flhD* leader, predicted by MFOLD webserver (Zuker, 2003), includes an upstream domain (nt 1-108) and an extensive inhibitory stem containing two internal loops (L I-II) and an internal hairpin (H I) (Fig. 3.8C). The downstream portion in this model is the same as the previous structural prediction using only the inhibitory stem (nt 109-206); it is also consistent with the functional studies suggesting that McaS sRNA binding opens up the secondary structure in the inhibitory stem (Thomason et al., 2012). A much less base-paired structure was reported by RNase footprinting for a 3'-truncation (nt 1-214) (Yakhnin et al., 2013). Nevertheless, the inhibitory stem structure in my model is highly supported by the increased SHAPE modification level at L II, H I, and the 3' portion of L I (Fig. 3.8B and C).

The conformation of the upstream domain was less well-determined. The upstream domain was poorly modified, except for one hyper-reactive A-nucleotide (nt 108) at the immediate 5' side of the inhibitory stem (Fig. 3.8B and C). This apparent lack of modification may result from a highly ordered structure that was resistant to SHAPE modification and that impeded reverse transcriptase during primer extension, leading to reduced signal and increased background. Another possibility is that the RNA sample was heterogeneous, although the RNA was purified from a single band in denaturing gel electrophoresis and was not frozen and thawed for more than 5 times. For future studies, I suggest preparing the RNA by applying the native gel purification protocol for making *rpoS* SAXS samples and keeping the RNA at 4 ºC for immediate usage within 1 week (Chapter 4 Materials and Methods).

My SHAPE results showed that the *flhD* leader directly binds Hfq distal face, because the (AAN) motif showed the signature hyperactive residues observed for the *rpoS* leader (Fig. 3.8A). However, based on my results, no other sites contact Hfq in a manner similar to the *rpoS* U_5 motif. The only hyperactive residue was nt 108, which was more likely caused by a distorted backbone conformation than direct Hfq binding. Coincidentally, Hfq did not remodel *flhD* secondary structure appreciably even at 10-fold higher molar concentration (Fig. 3.8D, E and F), consistent with the previous conclusion that a U_5 interaction was required for an Hfq-induced structural change.

Figure 3.8. SHAPE footprinting of *flhD* mRNA with Hfq. (*A*) Modification of the ribose 2′OH was quantified by primer extension and analyzed by sequencing gel electrophoresis (Materials and Methods). (*B*) To determine the overall secondary structure, traces from primer extensions covering different regions of the *flhD* RNA were overlapped. (*C*) *flhD* leader secondary structure consistent with experimental SHAPE modification data (colors). Loops regions showed higher SHAPE reactivity. Gray, no data. (*D*) SHAPE reactivity of 50 nM *flhD* RNA in complex with 333 nM Hfq hexamer (pink) or 500 nM Hfq (magenta) relative to *flhD* RNA alone. $(AAN)_4$ motif, nt. 169-177. NMIA modification was carried out at 37 ˚C (see Methods) and the extent of modification was measured by primer extension. Error bars represent ± S.D. for at least three independent experiments. (*E-F*) Schematic of SHAPE reactivity of each complex relative to free *flhD* RNA, from a histogram of the entire data set (not shown). Dark to light blue, protected, 0.46-0.54 (~1.4%) and 0.54-0.62 (~4.6%); black line, unchanged (~89.5%); light to dark red, enhanced, 1.42-1.98 (~3.4%), 1.98-3.02 (~0.8%), 3.02-3.26 (~0.4%); grey line, regions with no SHAPE data.


A
Sequence
SHAPE
- + + + NMIA
P A G C U - - + + Hfq
(AAN)4
B
20
16
12
8
4
0
SHAPE reactivity
nt 108
37 57 77 97 117 137 157 177
Nu #
C
flhD only
Upstream domain
50
100
5'
3'
200
L I
150
L II
(AAN)4
H I
D
3.5
3
2.5
2
1.5
1
0.5
0
Relative SHAPE reactivity
nt 108
flhD·Hfqlow
flhD·Hfqhigh
(AAN)4
35 55 75 95 115 135 155 175 195
Nu#
E
flhD·Hfqlow
nt 108
5'
3'
(AAN)4
F
flhD·Hfqhigh
nt 108
5'
3'
(AAN)4
-
+

Hfq facilitates post-transcriptional annealing of *flhD* and ArcZ56 RNAs

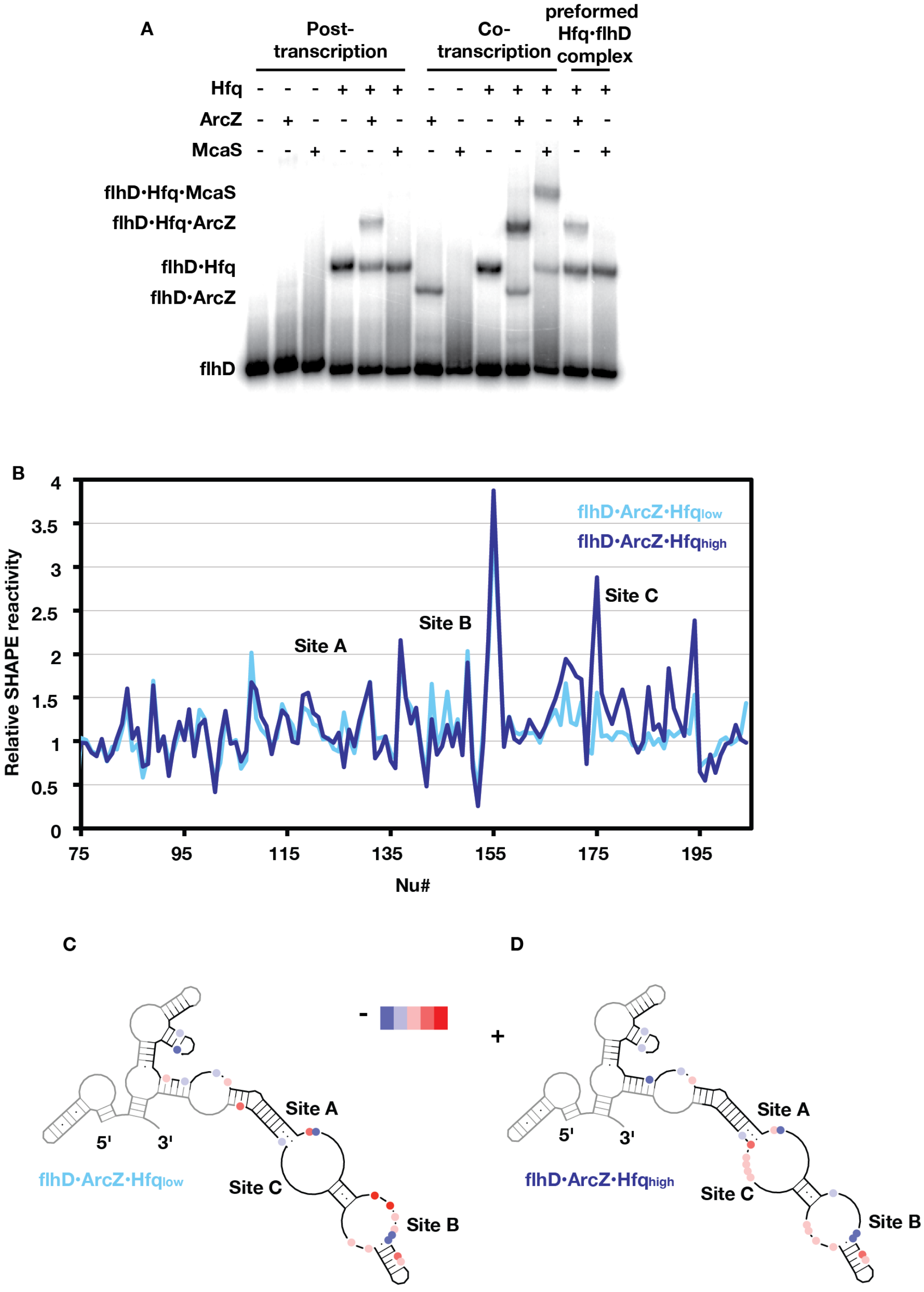
If U-rich motifs such as the one in the *rpoS* leader are important for efficient Hfq activity, the lack of a similar motif in the *flhD* leader may impair Hfq's ability to facilitate sRNA annealing. I tested *in vitro* annealing of *flhD* leader and two sRNAs that have been shown to directly base pair with *flhD* RNA *in vivo* (De Lay and Gottesman, 2012). In the McaS sRNA that positively regulates *flhD* translation, two discrete regions of McaS are predicted to base pair with sites A and B in the *flhD* leader with poor complementarity. A single 5' site in the ArcZ56 sRNA binds either site B or site C in *flhD* to inhibit ribosome binding (Fig. 3.7A) (De Lay and Gottesman, 2012).

Surprisingly, neither sRNA annealed with the folded *flhD* leader after transcription *in vitro* in the absence of Hfq (lane 1-3, Fig. 3.9A). In fact, base pairing between McaS and *flhD* RNAs was so poor that it did not occur even in the presence of 0.5 µM Hfq (lane 1, 3-4, and 6, Fig. 3.9A). By contrast, ~20% *flhD* formed a ternary complex with ArcZ56 and Hfq (lane 1, 4-5, Fig. 3.9A). RNA annealing occurred in the ternary complex because addition of poly(U) oligonucleotides, a Hfq binding competitor, cycled *flhD*•ArcZ56 RNA duplex off Hfq (data not shown). This moderate annealing enhancement is likely due to Hfq bringing the RNA strands into close proximity, because Hfq did not unwind *flhD* secondary structure to be captured by SHAPE (Fig. 3.8 E and F).

To test if the *flhD* RNA secondary structure hinders sRNA entry, I transcribed the *flhD* leader in the presence of ^{32}P-ATP, unlabeled sRNAs and Hfq, so that base pairing can occur before the *flhD* inhibitory stem has a chance to fold. ~10% *flhD* RNA annealed with ArcZ56 sRNA without Hfq; an additional 30% formed a ternary complex in the presence of Hfq (lane 1, 7, and 10, Fig. 3.9A). These data confirmed that unstructured *flhD* transcript not only annealed with ArcZ56 RNA alone but also promoted complex formation with Hfq and ArcZ56 RNA.

The *rpoS* leader possesses a single DsrA binding site that was less modified by NMIA in the ternary complex (Fig. 3.1D), suggesting that Hfq increased *rpoS* and DsrA annealing affinity and complex stability (Fig. 3.4C and Table 3.1) (Soper et al., 2010). Interestingly, two sites in the *flhD* leader can base-pair with the same sequence in ArcZ56 sRNA (site B and C, Fig. A) (De Lay and Gottesman, 2012), raising the possibility that Hfq may affect which of the two sites ArcZ56 occupies. I repeated SHAPE footprinting to investigate the structure of the *flhD* -ArcZ complex under different Hfq concentrations (Fig. 3.9B - D). Intriguingly, both site B and C showed 2- to 4-fold higher modification level in the ternary complex; suggesting that ArcZ binding may affect SHAPE modification (site B, Fig. 3.9C and site C, Fig. 3.9D). For instance, ArcZ binding at one site may open up local conformation to make the other site more accessible for NMIA. Moreover, varying the Hfq concentration shifted the modification enhancement from one site to the other (Fig. 3.9C and D), implying that Hfq may alter the structural details of *flhD*•ArcZ complex. This evidence is far from conclusive, however, due to the weak overall signals of SHAPE footprinting and subtle differences between data sets. Future experimental optimization is necessary for accurate structural probing.

Figure 3.9. Hfq facilitates post-transcriptional annealing of *flhD* and ArcZ56 RNAs. (*A*) Binding of ^{32}P-labeled *flhD* RNA with Hfq, ArcZ56 sRNA and McaS sRNA either after or during transcription (described in text and methods). (*B*) SHAPE reactivity of 50 nM *flhD* RNA in complex with 50 nM ArcZ56 sRNA and 333 nM Hfq hexamer (pink) or 500 nM Hfq hexamer (magenta) relative to *flhD* RNA alone. $(AAN)_4$ motif, nt. 169-177; site B, nt. 138-54; site C, nt. 178-194. NMIA modification was carried out at 37 ˚C (see Methods) and the extent of modification was measured by primer extension. (*C*-*D*) Schematic of SHAPE reactivity of each complex relative to free *rpoS* RNA as described in Fig. 3.8. Red circles, nucleotides with enhanced SHAPE reactivity; blue circles, nucleotides with reduced SHAPE reactivity; black line, regions with unchanged SHAPE reactivity; grey line, regions with no SHAPE data.


A
Post-transcription
Co-transcription
preformed Hfq·flhD complex
Hfq - - - + + + - - + + + + +
ArcZ - + - - + - + - - + - + -
McaS - - + - - + - + - - + - +
flhD·Hfq·McaS
flhD·Hfq·ArcZ
flhD·Hfq
flhD·ArcZ
flhD
B
Relative SHAPE reactivity
0
0.5
1
1.5
2
2.5
3
3.5
4
75
95
115
135
155
175
195
Nu#
flhD·ArcZ·Hfq$_{low}$
flhD·ArcZ·Hfq$_{high}$
Site A
Site B
Site C
C
-
+
5'
3'
Site A
Site B
Site C
flhD·ArcZ·Hfq$_{low}$
D
5'
3'
Site A
Site B
Site C
flhD·ArcZ·Hfq$_{high}$

Hfq facilitates co-transcription annealing of *flhD* leader and McaS sRNA

McaS sRNA only anneals with the *flhD* leader in the presence of Hfq during transcription (lane 1, 8-9, and 11, Fig. 3.9). I transcribed *flhD* RNA in the presence of ^{32}P-labeled McaS sRNA with or without wt Hfq and Hfq mutants, so that I was able to measure how fast McaS anneals with the nascent *flhD* transcript (Fig. 3.10 and Table 3.2). Under these conditions, the amount of flhDC rapidly exceeds that of McaS, so that all of the McaS sRNA should be able to bind if the sRNA-mRNA complex is stable.

Without Hfq, only 5% of McaS sRNA was able to bind *flhD* RNA after 60 min (Fig. 3.10A and grey trace, Fig. 3.10D). As expected, Hfq drastically improved RNA annealing by forming stable ternary complex in the gel (Fig. 3.10B). In fact, complex formation occurred at a rate constant of 0.28 ± 0.05 min^{-1} and reached a maximum of ~40% bound within 15 min (red trace, Fig. 3.10D). These results suggested that Hfq stabilizes the McaS-*flhD* mRNA complex and also restructures McaS sRNA to make it more competent to base pair with *flhD* mRNA. This restructuring activity was further confirmed by RNase footprinting experiments discussed in Chapter 5.

McaS was proposed to interact with both distal and proximal faces of Hfq (Zhang et al., 2013). I found that Hfq's activity was significantly disrupted by a single mutation on the distal face (Y25D), lateral surface (R16A), or proximal face (K56A), as well as truncation of the disordered C-terminal peptide (Hfq65) (purple, orange, green, and blue traces, Fig. 3.10D). All the Hfq variants reduced the annealing rate by 80-90% compared to wt Hfq (Table 3.2). The K56A mutation was the most deleterious, in that the annealing reaction appeared almost identical to the no Hfq control reaction (Fig. 3.10C). This proximal face dependency suggested that Hfq mainly acts by binding and remodeling

McaS sRNA to facilitate annealing. As a comparison, the *flhD*•ArcZ56 annealing reaction showed that WT Hfq increased the fast-annealing population from 15% to 80% without improving the annealing rate (Fig. 3.10E and Table 3.2), consistent with my previous conclusion that Hfq brings ArcZ56 into proximity with *flhD* RNA without chaperoning activity. Moreover, all the Hfq variants showed a similar ability to facilitate *flhD*•ArcZ56 annealing as WT Hfq, probably because all the mutants still bind the RNAs, which will be further discussed in Chapter 5.

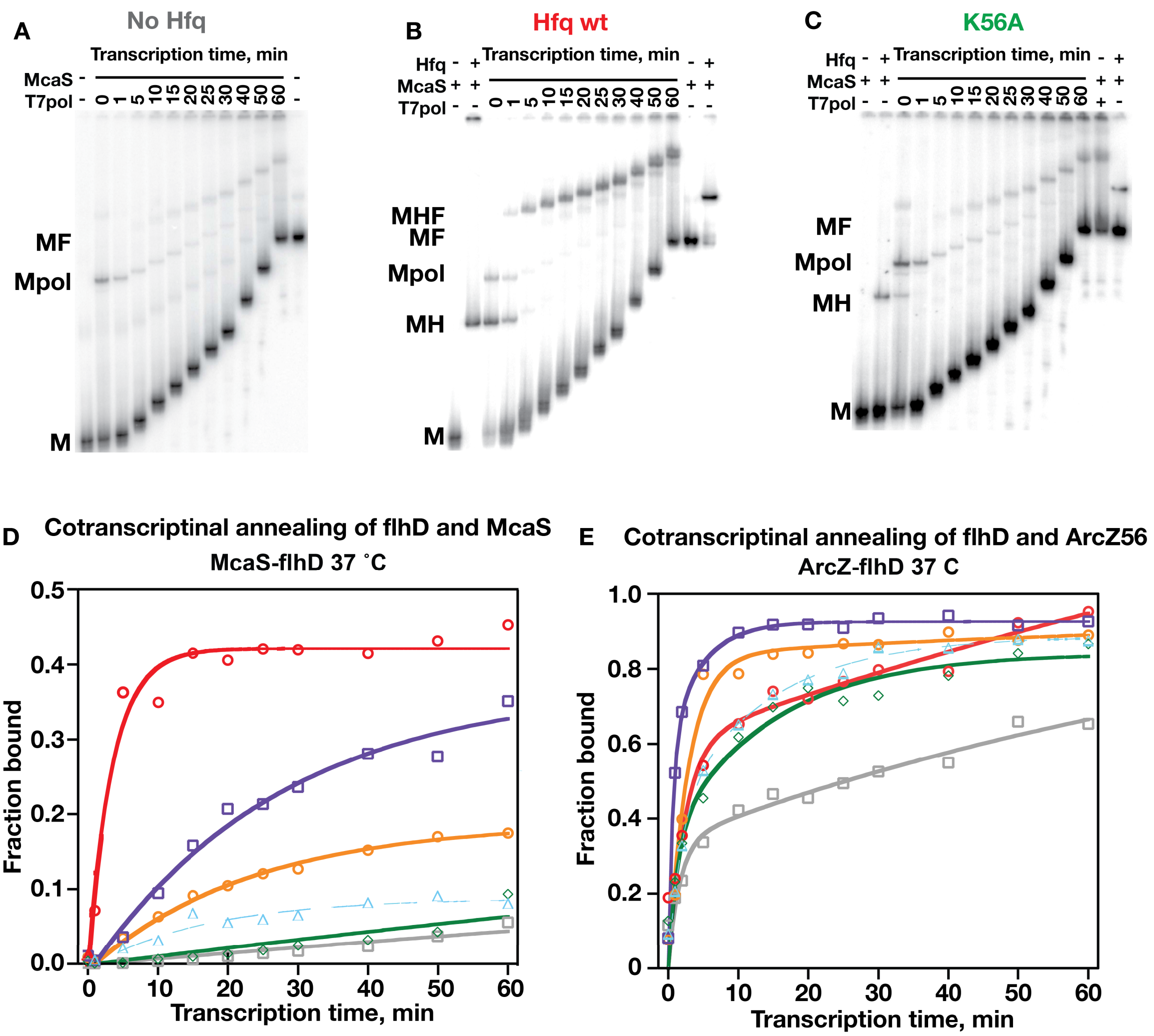


Figure 3.10. Hfq facilitates co-transcriptional annealing of *flhD* and sRNAs. Gel images of cotranscriptional annealing kinetics of *flhD* mRNA and McaS sRNA in the absence of Hfq (*A*) or presence of wt Hfq (*B*), Hfq K56A (*C*) and other variants (not shown). Fraction of McaS (*D*) and fraction of ArcZ56 (*E*) bound *flhD* leader were fit with a single or double exponential rate equation (Chapter 2 Material and Methods, eq. 2.3 - 2.4; Materials and Methods, eq. 3.1). Reactions contained no Hfq (grey), Hfq wt (red), R16A (orange), Y25D (purple), K56A (green), and Hfq65 (blue) (Materials and Methods).

Table 3.2. Summary of co-transcriptional annealing rate of *flhD* with McaS and ArcZ56 sRNAs

	McaS annealing rate		ArcZ56 annealing rate	
	Fraction	Rate (min^{-1})	Fraction	Rate (min^{-1})
No Hfq	0.42	0.0019 ± 0.0001	0.33 ± 0.07 (fast)	0.61 ± 0.27
			0.73 ± 1.27 (slow)	0.01 ± 0.03
Wt Hfq	0.42 ± 0.01	0.28 ± 0.05	0.60 ± 0.13 (fast)	0.42 ± 0.17
			1.24 ± 7.62 (slow)	0.005 ± 0.04
Y25D	0.39 ± 0.04	0.03 ± 0.01	0.59 ± 0.17 (fast)	1.49 ± 0.65
			0.33 ± 0.17 (slow)	0.22 ± 0.11
R16A	0.19 ± 0.01	0.04 ± 0.005	0.84 ± 0.12 (fast)	0.35 ± 0.08
			0.11 ± 1.06 (slow)	0.01 ± 0.20
K56A	0.42	0.0028 ± 0.003	0.36 ± 0.15 (fast)	0.75 ± 0.50
			0.49 ± 0.13 (slow)	0.07 ± 0.03
Hfq65	0.09 ± 0.01	0.06 ± 0.02	0.41 ± 0.14 (fast)	0.53 ± 0.22
			0.48 ± 0.12 (slow)	0.07 ± 0.03

flhD leader was transcribed in the presence of P^{32}-labeled sRNA and Hfq at 37 ºC (Materials and Methods). Selective electrophoresis gels were shown in Fig. 3.10. Annealing reactions of ArcZ56 and flhD RNAs were fit with a double exponential rate equation (Chapter 2 Material and Methods, eq. 2.4). Reaction of McaS and flhD RNA in presence of wt Hfq was fit with a single exponential rate equation (Chapater 2 Material and Methods, eq. 2.3), while the other reactions were fit with a single exponential rate equation assuming a lag time of 1 min (Material and Methods, eq. 3.1).

CsrA reduced *flhD* annealing by forming complex with Hfq and McaS

My *in vitro* data suggested that McaS and ArcZ56 sRNAs did not anneal with *flhD* leader very well even with Hfq present; however, previous *in vivo* studies showed a strong translation regulation of *flhD* leader by those sRNAs (De Lay and Gottesman, 2012; Thomason et al., 2012). I asked whether another protein regulator is involved in translational regulation of the *flhD* RNA.

One candidate is CsrA, a homodimeric protein that controls translation of target genes by binding the GGA motifs in the mRNAs (Timmermans and Van Melderen, 2010). Interestingly, McaS sRNA has been shown to act in both CsrA- and Hfq-dependent regulatory pathways by directly binding to the two global protein regulators (Jorgensen et al., 2013), raising the possibility that Hfq and CsrA may concurrently act on McaS and its target mRNAs. In fact, Hfq binding to RsmY RNA stabilized the complex in which the RNA sequestered the CsrA homolog RsmA in *P. aeruginosa* (Sorger-Domenigg et al., 2007). To investigate whether and how CsrA and Hfq proteins concomitantly regulate *flhD*•McaS annealing, I first investigated post-transcriptional binding of the *flhD* leader to McaS sRNA, Hfq and CsrA (Fig. 3.11).

^{32}P-labeled *flhD* RNA was incubated with Hfq only, CsrA only, or both Hfq and CsrA in the absence or presence of McaS RNA for 1 hr; the complex formation was resolved by native polyacrylamide gel electrophoresis. Without McaS, not only did *flhD* bind Hfq or CsrA, respectively (lane 1-3, Fig. 3.11A) (De Lay and Gottesman, 2012; Yakhnin et al., 2013), it also formed a ternary complex with both proteins. McaS reduced the amount of *flhD*•protein complexes, although the same complexes were still observed (lane 7-8, Fig. 3.11A). This is consistent with the previous observation that noncomplementary RNAs act as competitors to cycle off existing RNAs from Hfq (Soper et al., 2011).

Since Hfq was able to facilitate co-transcriptional annealing of *flhD* and McaS RNAs, I next examined how CsrA affected the annealing reaction (Fig. 3.11B-D). CsrA by itself did not improve McaS•*flhD* duplex formation (Fig. 3.11B; grey and green trace, Fig. 3.11D), although it initially formed a weak complex with McaS that was competed off by synthesis of *flhD* within 1 min (Fig. 3.11B). Intriguingly, when both Hfq and CsrA were present in the annealing reaction, ~40% McaS remained in a stable Hfq•McaS•CsrA complex that slowly annealed with *flhD* RNA in 30 min (Fig. 3.11C), suggesting a reduced annealing rate for the sequestered McaS•Hfq (Fig. 3.11C and D). In fact, the annealing rate was $0.03 \pm 0.01\ \text{min}^{-1}$ for reaction containing both CsrA and Hfq as compared to $0.12 \pm 0.01\ \text{min}^{-1}$ for reaction containing Hfq only. The fraction bound was also reduced from ~25% to ~15%. One explanation is that CsrA stabilizes the McaS•Hfq complex and impairs restructuring of McaS RNA for effective *flhD* annealing. This explanation would be consistent with the observation that the CsrA homolog RsmA stabilized the ternary complex of RsmA•RsmY•Hfq to resist RNase E digestion (Sorger-Domenigg et al., 2007).

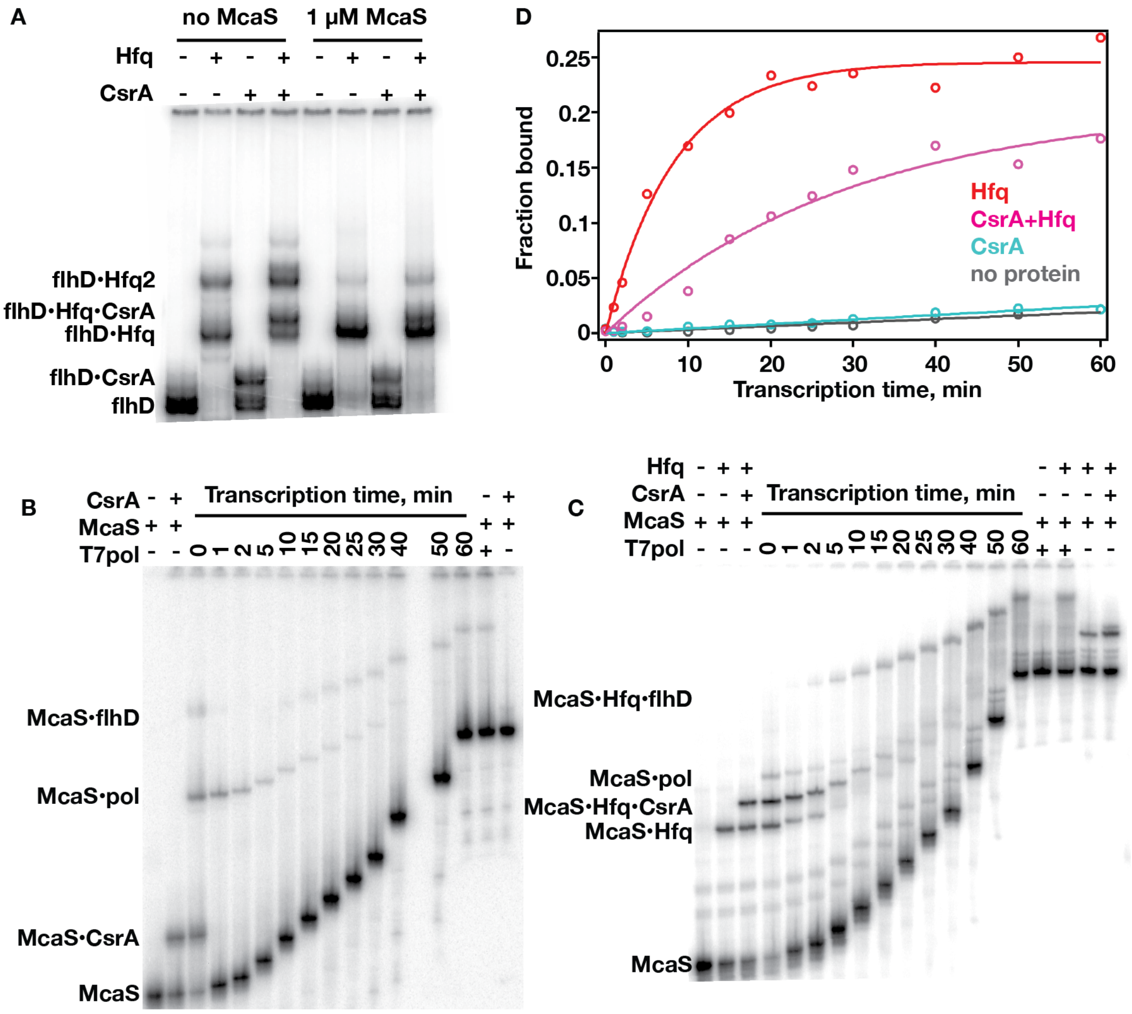


Figure 3.11. CsrA reduced Hfq's annealing activity. (*A*) Complex formation of 32P-labled flhD leader with CsrA, Hfq and McaS RNA. (*B*, *C*) Gel images of cotranscriptional annealing kinetics of flhD mRNA and McaS sRNA in the presence of CsrA without or with Hfq. (*D*) Fraction of McaS sRNA bound with flhD leader was fit with a single exponential rate equation (Chapter 2 Material and Methods, eq. 2.3). Reactions contained no Hfq (grey), CsrA only (green), CsrA and Hfq (magenta), and Hfq only (red) (Materials and Methods).

Discussion

While AAN sequences are known to recruit Hfq via its distal face (Link et al., 2009; Mikulecky et al., 2004; Soper and Woodson, 2008), here I find that the U_5 motif in *rpoS* also contributes to sRNA annealing by interacting with the Hfq lateral surface. This distorts the mRNA structure, making it more accessible to sRNAs (Soper et al., 2011). Multi-lateral Hfq interactions may be widespread among bacterial sRNA-mRNA pairs and important for regulation. The *fhlA* mRNA leader was proposed to contact both distal and proximal faces of Hfq based on competitive binding experiments (Salim and Feig, 2010). Hfq inhibits translation of *cirA* by binding to an upstream (AAN) motif and two U-rich patches close to the Shine-Dalgarno sequence (Salvail et al., 2013).

By contrast, my initial data indicate that Hfq only binds a (AAN) motif in the *flhD* leader and does not restructure this RNA significantly. This may be caused by the lack of U-rich motif that binds Hfq lateral surface to distort RNA conformation, or it may result from *flhD*'s highly base-paired secondary structure. This Hfq binding mode leads to different annealing mechanisms for *flhD*'s regulatory sRNAs ArcZ56 and McaS. In the case of ArcZ56, Hfq mainly brings the *flhD* leader into proximity to base pair with the two extensive complementary regions located in a relatively simple secondary structure. By contrast, I propose that Hfq bridges and restructures McaS RNA to overcome its poor complementarity with *flhD* leader. Moreover, *flhD*•McaS is the first reported mRNA•sRNA pair that only anneals during transcription *in vitro*, providing a potential system to study coupled transcription-translation events in bacterial cells. Further study is required to explore the implications of co-transcriptional annealing *in vitro* for Hfq's chaperone activities *in vivo*.

Finally, the global regulators CsrA and Hfq concurrently bind the *flhD* leader (De Lay and Gottesman, 2012; Yakhnin et al., 2013), McaS (Jorgensen et al., 2013), and RsmY sRNAs (Sorger-Domenigg et al., 2007). Simultaneous interactions with CsrA and Hfq proteins may occur for other RNAs that possess CsrA-binding GGA motifs and Hfq-binding AAN motifs. The ternary complex containing both proteins was more stable than RNA complexes with either protein alone, because this hyper-stable complex acts as a molecular sequester to repress RNA processing activity (Sorger-Domenigg et al., 2007). The observation applies to RsmY (Sorger-Domenigg et al., 2007) and McaS sRNAs, and may serve as a general mechanism that fine-tunes the crosstalk between CsrA- and Hfq-dependent pathways.

Material and Methods

SHAPE footprinting

E. coli Hfq protein, *rpoS301* RNA and DsrA sRNA were prepared as described in Chapter 2 (Peng et al., 2014). Reactions (10 µL) were assembled in annealing buffer (50 mM Tris HCl pH 7.5, 50 mM NaCl, 50 mM KCl, 50 mM NH_4Cl, 2% glycerol) and contained 50 nM *rpoS301* for *rpoS* alone, plus 200 nM DsrA and 333 nM Hfq hexamer for their respective complexes. Reactions were incubated at 25 °C for 2 hours to allow complete complex formation. SHAPE modification was carried out by adding 1 µL of 20 mM N-methylisatoic anhydride (Molecular Probes) dissolved in anhydrous DMSO (Molecular Probes) or 1 µL DMSO only, followed by incubation at 37 °C for 2 h (Wilkinson et al., 2006). The reaction was diluted with 189 µL H_2O, extracted with 200 µL phenol, 200 µL chloroform. The modified RNA was precipitated with ethanol and resuspended in 10 µL H_2O. Primer extension was carried out using ^{32}P-labeled primers as previously described (Peng et al., 2012).

SHAPE footprinting of the *flhD* leader was done as described above, except that each 10 µL reaction contained 50 nM *flhD* RNA in combination with 200 nM ArcZ56, 333 nM Hfq hexamer (for Hfq_{low}) and 500 nM Hfq hexamer (for Hfq_{high}) for their respective complexes.

SHAPE data analysis

cDNA fragments (sequencing gel bands) were quantified using SAFA (Das et al., 2005)that was modified to normalize band intensity in each lane with respect to reference bands whose intensity did not change in the experiment (Behrouzi, 2012). Corrections to the loading error ranged from 0.95 – 1.06. The relative SHAPE reactivity of *rpoS* RNA complexes versus *rpoS* RNA alone was calculated from the ratio of adjusted band intensities, with values ranging from 0.15 to 8.88. Outliers with values >30 were eliminated from further analysis. These usually occurred next to very strong bands and in the less resolved portion of the gel. SHAPE reactivities from at least three independent experiments were averaged.

Relative SHAPE reactivities for *rpoS* leader were categorized as unchanged (0.728-1.61, ~87% of data), protected (0.14-0.58, ~3%), and enhanced (1.76-8.9, ~10%) (Fig. 3.2C). Relative SHAPE reactivities for the *flhD* leader were categorized as unchanged (0.62-1.42, ~89% of data), protected (0.46-0.62, ~6%), and enhanced (1.42-3.18, ~5%). RNA secondary structure schematics were drawn with XRNA (developed by Noller lab at UCSC, http://rna.ucsc.edu/rnacenter/xrna/xrna.html).

β-galactosidase assays and in vitro gel mobility shift assays

The *rpoS* mutants described in the text (ΔU_5, U_5SS, and U_5UL) were generated by inverse PCR as described in Chapter 2 (sequences listed in Appendix II) (Peng et al., 2014). *E. coli* strains and β-galactosidase assays of *rpoS::lacZ* expression were performed as described in Chapter 2 (Peng et al., 2014).

Gel mobility shift assays

Preparation of ^{32}P-labeled RNA and gel mobility shift assays were done as described in Chapter 2 (Peng et al., 2014; Soper and Woodson, 2008). For DsrA•*rpoS* annealing, ~70 nM labeled *rpoS301* in annealing buffer was mixed with DsrA and Hfq as indicated in the text. [Hfq] was shown for monomers. For equilibrium experiments, reactions were incubated at 25 °C for 2 hours and loaded on a native 6% polyacrylamide gel in 1X THEM2 buffer (66 mM Hepes, 34 mM Tris, 0.1 mM EDTA, and 2 mM $MgCl_2$). For kinetics experiments, reactions were incubated at 25 °C for different times before loading.

For post-transcriptional equilibrium experiment of *flhD* RNA, the reaction was set up as described above, except that the native polyacrylamide gel electrophoresis was done in 1X TBEM3 buffer (89 mM Tris, 89 mM boric acids, 2 mM EDTA, and 3 mM $MgCl_2$).

Co-transcriptional binding

For co-transcriptional equilibrium experiments, 8 µl transcription reaction mixture was prepared, containing 0.5 µCi/µl α-^{32}P-ATP, 0.125 mM ATP, 0.5 mM GTP/CTP/UTP, 12.5 ng/µl DNA template in 1X T7 RNAP buffer (40 mM Tris-HCl pH 8.0, 15 mM $MgCl_2$, 5 mM DTT, and 2 mM spermidine). The DNA template for transcribing *flhD* RNA was generated by PCR (primers were listed in Appendix II). 1 µl ArcZ sRNA at various concentrations (Fig. 3.9B) was incubated at 25 ºC with 1 µl of 5 µM Hfq monomers or Hfq buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 250 mM NH_4Cl, and 10% glycerol) and added to the transcription reaction mixture. 1 µl T7 RNA polymerase was added to the 10 µl reaction and the reaction was incubated at 37 ºC for 30 min. A 2 µl aliquot was loaded on a native 6% polyacrylamide gel in 1X TBEM3 buffer.

For co-transcriptional kinetics experiments, 40 µl transcription reaction contained 0.5 mM NTPs, 12.5 ng/ul DNA template, ~50 nM ^{32}P-labeled sRNA with or without 0.6 µM Hfq monomers or 0.25 µM CsrA monomers in 1X T7 RNAP buffer. The mixture was incubated at 37 ºC for 10 min before adding 1 µl T7 RNA polymerase to start transcription. A 2 µl aliquot was loaded in the native polyacrylamide gel after different times as indicated in Fig. 3.10-3.11.

The fractions of bound *flhD* were quantified as previously described (Lease and Woodson, 2004) and were fit with either a single or double exponential rate equation (KaleidaGraph, Synergy Software) as described in Chapter 2 Materials and Methods eq. 2.3 and 2.4. Some reactions were fit with a single exponential rate equation assuming lag time of 1 min as indicated in Table 3.2.

$$f_{RD} = f(1-\exp(-k_{obs1}(t-1))) \tag{3.1}$$

in which f is the fraction of the reaction that followed the corresponding annealing rate, t is the annealing time, k_{obs1} is the apparent annealing rate.

Chapter 4 Folded solution structure of the *rpoS* mRNA leader in complex with Hfq

Introduction

RpoS is a σ factor that regulates transcription of stress responses genes (Battesti et al., 2011). Translation of *rpoS* mRNA is activated by sRNAs that base pair with the sRNA annealing site to release the ribosome binding site from the inhibitory stem (Majdalani et al., 1998; Majdalani et al., 2001; Mandin and Gottesman, 2010). sRNA-mRNA annealing is mediated by bacterial RNA chaperone Hfq, which binds to a $(AAN)_4$ motif in the *rpoS* leader on its distal surface (Soper et al., 2010; Soper and Woodson, 2008). Interestingly, the A-rich motif is ~60 nt away from the inhibitory stem (Chapter 2 Fig. 2.1), where Hfq recruits the sRNA and unwinds the local secondary structure near the translation starting site (Hammerle et al., 2013; Soper et al., 2011). Moreover, the distance and orientation of the upstream binding site and the downstream action site is important for optimal Hfq function (Peng et al., 2014), raising the possibility that Hfq may fold the *rpoS* leader to coordinate the different functional elements. I discovered that Hfq lateral rim also directly binds to a U_5 motif downstream of the sRNA annealing site, showing that Hfq simultaneously contact two sites that are ~200 nt apart in the *rpoS* leader (Chapter 3). All the evidence suggests that Hfq may alter the tertiary conformation of the *rpoS* leader to remodel the inhibitory stem for sRNA annealing and translation initiation.

To investigate that hypothesis, I used small angle X-ray scattering (SAXS) to measure the conformational change of the *rpoS* leader upon Hfq binding. SAXS captures the scattering profiles of X-rays traveling through macromolecules in solution at very low angles (typically 0.1 – 10º), which can be applied to a wide range of solution conditions, temperatures, particle sizes, and compositions (Hura et al., 2009). It is a snapshot of a structural ensemble that requires careful interpretation to obtain reliable information regarding the size, shape, and flexibility of the particle (Glatter and Kratky, 1982). For instance, the radius-of-gyration (R_g) describes the particle's distribution of mass around the center of gravity (Blanchet and Svergun, 2013); changes in R_g have been used to measure folding pathways of several large RNAs (Lipfert et al., 2008; Perez-Salas et al., 2004). Furthermore, computational programs, such as DAMMIN and DAMMIF (Svergun, 1999), have been used to reconstruct *ab initio* models of large RNA-protein particles at low resolution, including Hfq in complex with RprA and OxyS sRNAs (Henderson et al., 2013). Rigid body modeling approach, given the atomic structures of individual components, has been developed to generate quaternary conformations of proteins (Petoukhov and Svergun, 2005). In the case of flexible molecules, ensemble modeling strategies have been employed to generate a library of conformers to fit the SAXS data (Rambo and Tainer, 2013).

Here, I show that Hfq enables sRNA regulation by folding the *rpoS* mRNA leader into a specific tertiary structure that partially unwinds the inhibitory stem and poises Hfq to bring both RNAs together. Small angle X-ray scattering (SAXS) revealed a more compact conformation of the *rpoS* leader upon Hfq binding. Three-dimensional models of the *rpoS*•Hfq complex refined against the SAXS data show that the two domains of the *rpoS* mRNA wrap around the Hfq hexamer, placing the inhibitory stem over the arginine patch and adjacent to the sRNA binding sites on the rim and proximal face. These results demonstrate that multiple RNA binding surfaces on Hfq enable the protein to distort the structure of the *rpoS* mRNA, poising the complex for sRNA entry and translation.

Results

Hfq folds *rpoS* mRNA

If the distal face of Hfq binds the upstream $(AAN)_4$ motif while the lateral rim interacts with the downstream U_5 motif, Hfq binding likely alters the tertiary conformation of the *rpoS* mRNA. To test that hypothesis, I used SAXS to compare the global shape of free *rpoS* RNA and the *rpoS*•Hfq complex in solution (Fig. 4.1). *rpoS* RNA and Hfq proteins were prepared under non-denaturing conditions and dialyzed against SAXS buffer before assembling the reactions to ensure accurate buffer subtraction (Materials and Methods). The homogeneity of the RNA sample was measured by dynamic light scattering; a small amount of *rpoS* dimer (11% of dynamic light scattering; <1% number density) was detected after samples were stored at -80 ˚C (Fig. 4.2). This RNA dimerization occurred at high concentration and may be aggravated by the freeze-thaw cycle that was avoided for the SAXS measurements. Consequently, the freshly prepared sample used for SAXS likely contained less RNA dimer than shown in Fig. 4.2. Nevertheless, this minor population of *rpoS* dimers (if present) was not readily detected in the scattering data.

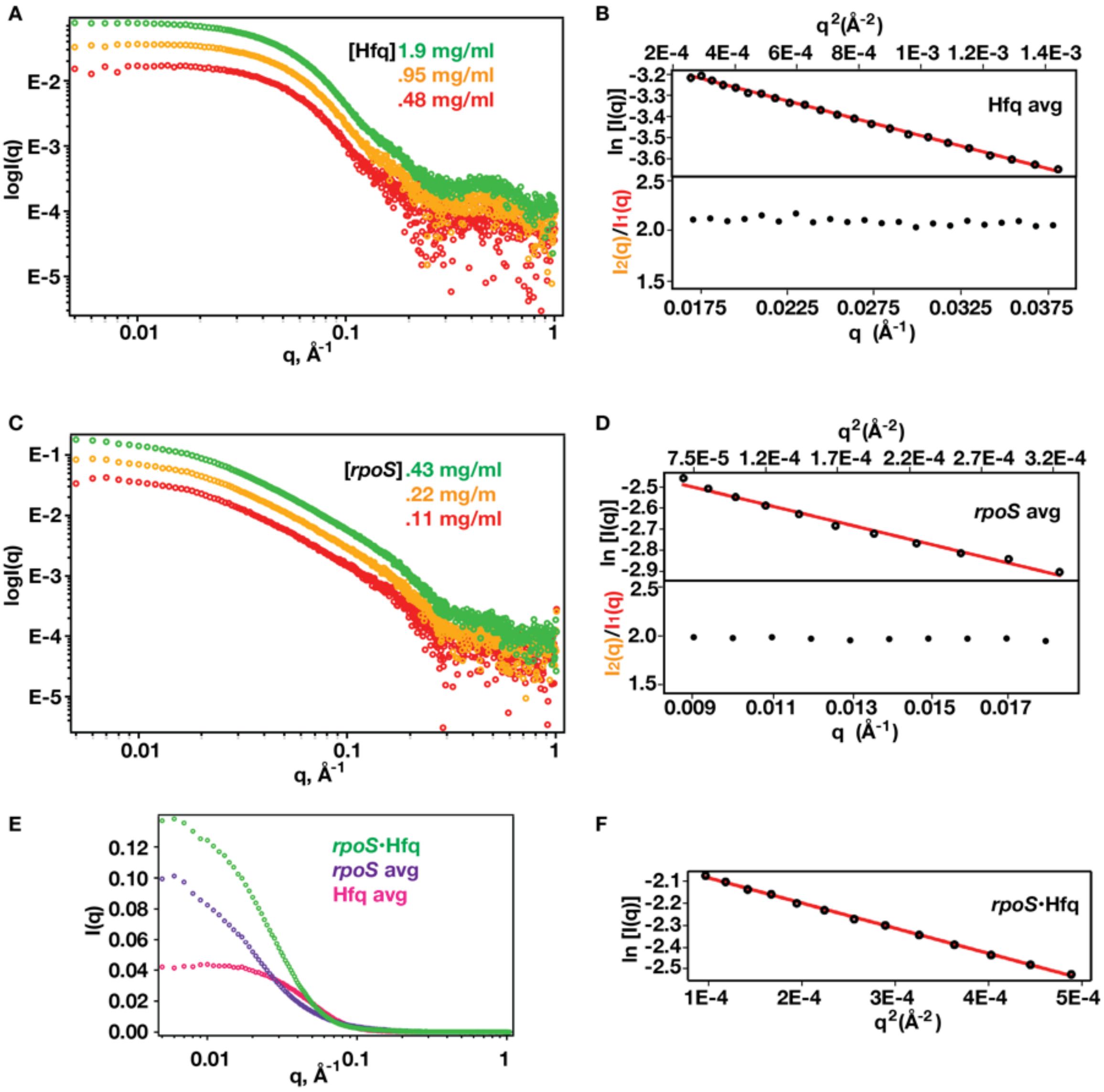


Figure 4.1. Solution scattering of Hfq and *rpoS* mRNA. SAXS scattering profiles of (*A*,*B*) full-length Hfq and (*C*,*D*) *rpoS301* RNA at three concentrations (colored green, orange, and red) in solution. (*B*,*D*) Top panels: Guinier plot of averaged data for free Hfq and free *rpoS*301 RNA. Bottom panels: The ratio of scattering intensity from two Hfq or RNA concentrations remained constant at the Guinier region, confirming the absence of inter-particle interactions. (*E*) The averaged scattering curves for Hfq (magenta), *rpoS* RNA (purple) and 1:1 *rpoS*•Hfq complex (green) were used to calculate R_g from the Guinier region (33.6 ± 0.5 Å, 68.1 ± 1.6 Å, and 58.0 ± 1.0 Å, respectively). (*F*) Guinier plot of 1:1 rpoS•Hfq complex.

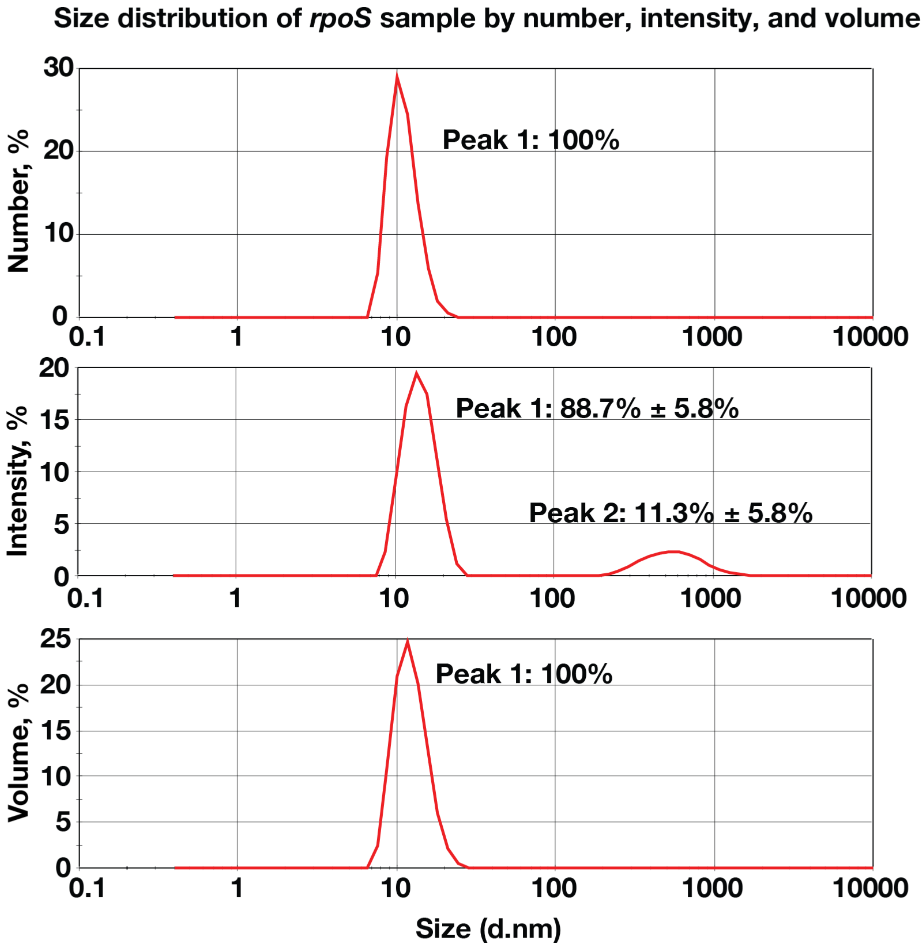


Figure 4.2. Dynamic light scattering of free *rpoS* RNA. Dynamic light scattering measurements were performed retrospectively on RNA samples prepared from stock solutions used to prepare SAXS samples. Average number of molecules, scattering intensity, and molecular volume is shown as a function of particle size, with the average peak intensities ± S.D. of three measurements. The minor peak of scattering intensity corresponds to an *rpoS* RNA dimer. These samples were frozen and thawed several times before DLS. It is likely that the original SAXS examples that were stored at 4 ˚C contained less dimer than shown here. Moreover, RNA sample was diluted for various reactions (Materials and Methods) that may alleviate dimerization.

The scattering profile of free Hfq protein (Fig. 4.1A and B) was consistent with its known structure as previously reported (Beich-Frandsen et al., 2011; Henderson et al., 2013). The scattering profile of the free *rpoS* mRNA (284-nt) revealed an extended structure with radius of gyration (R_g) = 68.1 ± 0.6 Å (Fig. 4.1C and D). A dimensionless Kratky plot of the scattering intensity exhibited a plateau at higher momentum transfer (*q*) indicating an extended or flexible conformation (Fig. 4.3A, grey symbols).

Addition of Hfq to *rpoS* sample at molar ratios from 1:0.5 to 1:3 RNA:Hfq_6 dramatically changed the shape of the Kratky scattering curves, forming an almost symmetric and bell-shaped curve characteristic of globular particles (Fig. 4.3A). Interestingly, the scattering curve shifts with Hfq concentration. At Hfq:*rpoS* molar concentration = 0.5, the scattering curve resembles free *rpoS* with a elongated tail at the high q region (Fig. 4.3 lightest pink symbols). Increasing Hfq concentration up to 2:1 molar ratio led to a more symmetric curve indicating a more compact conformation (Fig. 4.3 pink to bright red symbols). Further Hfq titration incrementally increased the amplitude of the curve peaks, suggesting a formation of higher order complexes (Fig. 4.3 dark red and brown symbols). This change in shape corresponded with a dip in R_g values at 1:1 molar ratio (Fig. 4.3B, solid circles) despite the gradually increased mass of the *rpoS*•Hfq complex (Fig. 4.3B, open circles). The smallest R_g value of 58 ± 1 Å was reached at 1:1 r*poS*:Hfq_6 (Fig. 4.1E and F), at which concentration 95% of the RNA is expected to be bound with Hfq (Peng et al., 2014). This was further confirmed by a native gel mobility shift analysis of recovered SAXS samples showing that 1:1 molar concentration yielded the greatest amount of complex (~87%) (Fig. 4.3D).

The change in the scattering profile cannot be explained by scattering from the protein alone, as Hfq has a much smaller X-ray scattering contrast than the RNA (Fig. 4.1E). Instead, I inferred that the flexible *rpoS* leader must adopt a more compact tertiary structure when bound to Hfq. This compact structure is stabilized by interactions between *rpoS* mRNA and the lateral surface of Hfq, because the Hfq:R16A mutant formed a more extended complex with *rpoS* mRNA than did WT Hfq (Fig. 4.3C, gold symbols).

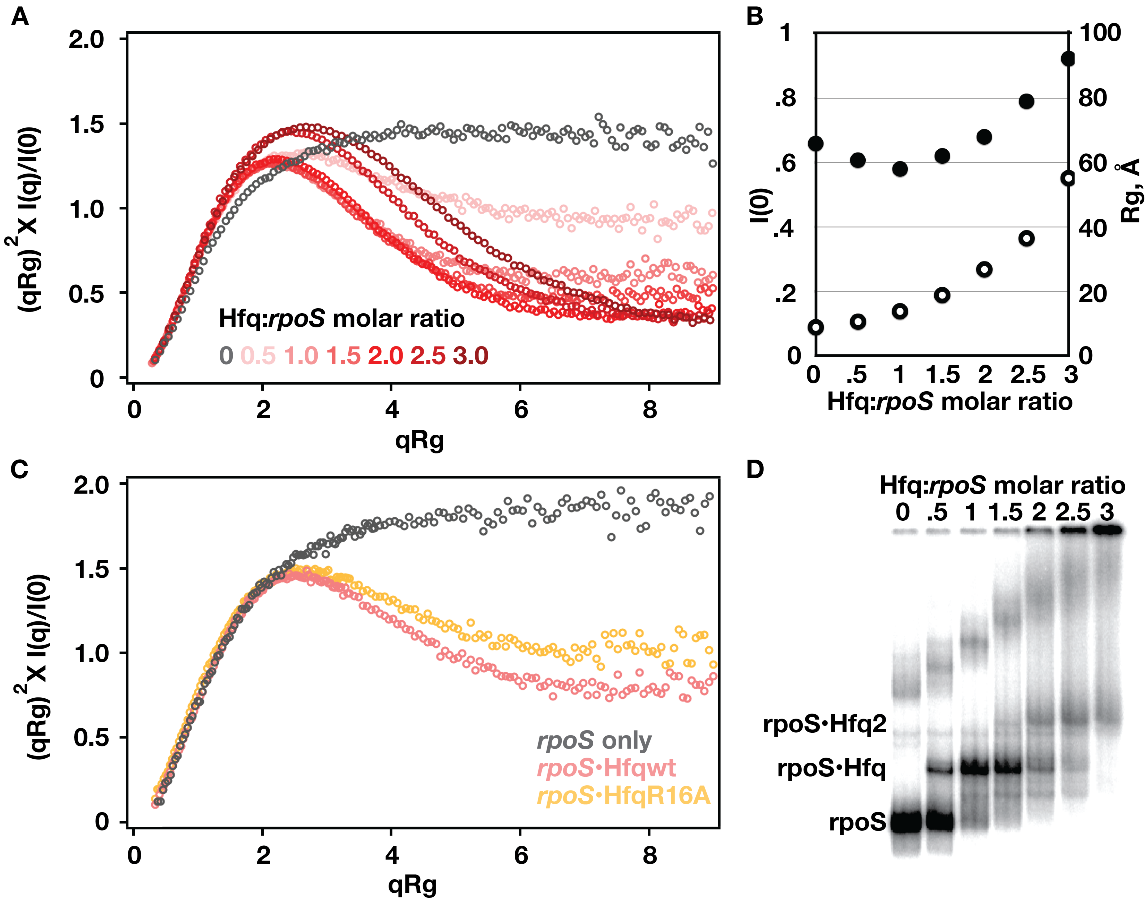


Figure 4.3. SAXS of *rpoS*•Hfq complexes reveals a compact structure. (*A*) Dimensionless Kratky plot (Durand et al., 2010; Receveur-Brechot and Durand, 2012) of SAXS profile for *rpoS* RNA alone (grey) and *rpoS*•Hfq complexes at increasing protein:RNA ratios (pink to red). Bell shaped curves indicate compact structures. See Fig. 4.1 and Table 4.1 for further data. (*B*) Hfq binding increased I(0) (open circles) and decreased R_g (solid circles) compared to free *rpoS* RNA. At 1:1 mole ratio, ~95% of RNA is bound to Hfq. Correction for scattering from the free RNA and protein reduces the experimental R_g of the complex by ~1 Å. (*C*) Kratky plots of *rpoS* RNA alone (grey), with 1:1 WT Hfq (pink) and with 1:1 Hfq:R16A (gold). (*D*) Native polyacrylamide gel electrophoresis measured the fraction of *rpoS*•Hfq complexes in the SAXS samples recovered after radiation. ~87% *rpoS* RNA was bound to Hfq at 1:1 molar ratio as quantified from the gel image. Correction for scattering from the free RNA and protein reduces the experimental R_g of the complex by ~1-2 Å, within the error of the models.

Ab initio modeling

If Hfq folded *rpoS* RNA in solution, I expected to see a shape change in the molecular envelope derived from SAXS scattering. I first used the SAXS results to generate *ab initio* model of the free *rpoS* mRNA and the *rpoS*•Hfq complex. This modeling procedure generates structural decoys by randomly placing spherical particles, referred to as dummy atoms, in a defined 3D searching space; and selects a configuration that results in a calculated SAXS scattering profile best matched with the input experimental data (Svergun, 1999). The input SAXS data of free *rpoS* and Hfq were averaged from SAXS profiles at three different concentrations (Fig. 4.1A, C and E), which displayed linearity in the Guinier region and were free of inter-particle interactions (Fig. 4.1B and D). The SAXS data used for modeling *rpoS*•Hfq complex was generated from the sample at Hfq:*rpoS* molar ratio = 1:1 (Fig. 4.1E and F), because it showed the maximum 1:1 complex formation (Fig. 4.3A, B, and D).

Molecular envelopes calculated *ab initio* from the SAXS data revealed an elongated "L" for the free *rpoS* RNA (Fig. 4.4A), which curled inward when Hfq was present (Fig. 4.4B). It is conceivable that Hfq binding may fold the RNA into a compact conformation by "holding" the upstream and downstream domains. Nevertheless, many structural features were lost when these envelopes were averaged, presumably because the RNA is flexible and poorly constrained by the scattering curves. In addition, Hfq was nearly invisible in the molecular envelopes, owing to its lower scattering contrast relative to the RNA. Therefore, I used rigid-body methods to build atomistic models of the *rpoS*•Hfq complex using the information available from SAXS, crystal structures, and biochemical footprinting. Although these data cannot specify the conformation of individual residues, I obtained low-resolution models that were consistent with all the available data and that suggested how Hfq enables sRNA regulation of *rpoS* translation.

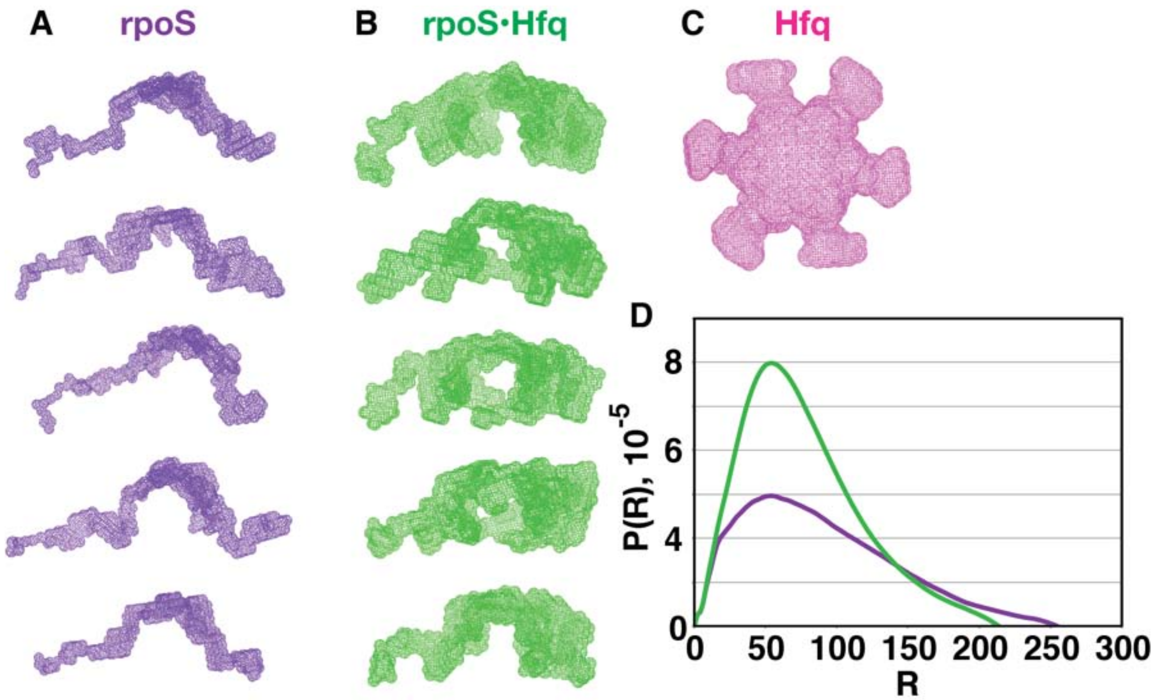


Figure 4.4. ***Ab initio*** **models of** ***rpoS*** **RNA, the** ***rpoS*****•Hfq complex, and Hfq protein.** (*A*,*B*) Five ab initio structures predicted by DAMMIF for (*A*) free *rpoS* RNA and (*B*) the *rpoS*•Hfq complex (1:1 molar ratio). The *rpoS* envelopes adopt an extended L-shaped conformation, consistent with all-atom model (Fig. 4.5 and 4.7). Hfq binding folds *rpoS* RNA into a more compact conformation. DAMMIF assumes the electron density is evenly distributed in the complex, and does not account for differential SAXS scattering intensity from Hfq protein and *rpoS* mRNA. As a result, the space occupied by Hfq protein appears empty in the molecular envelope. Averaging these *ab initio* models resulted in a significant loss of structural detail. (*C*) The averaged *ab initio* model of Hfq predicted by DAMAVER (V. V. Volkov and D. I. Svergun, 2003) is very similar to previous models based on SAXS data (Beich-Frandsen et al., 2011; Henderson et al., 2013), which assumed oblate P6 symmetry. Twenty DAMMIF bead models were averaged and DAMMIN was restarted to fit the experimental data. (*D*) *P(R)* distribution of free *rpoS* RNA (purple) and the *rpoS*•Hfq complex (green) showing the change in the average conformation of the particle. The mass of the RNA and Hfq particles were calculated from the molecular volume, Vc, using the method of Rambo and Tainer (Rambo and Tainer, 2013). Real space parameters for the scattering data are given in Table 4.1.

Table 4.1. SAXS parameters for Hfq protein and *rpoS* mRNA

		Guinier				Real space					Mass (Da) (1)	
Sample	(µM)	R_g (Å)	+/-	I(0)	+/-	R_g (Å)	+/-	I(0)	+/-	R_{max}	SAXS	nominal
Hfq	42.4	32.3	0.7	1.86E-02	4.6E-04	32.8	0.8	1.90E-02	3.0E-04	110	73,762	66,996
	84.8	33.7	0.9	3.95E-02	5.4E-04	34.0	0.4	3.95E-02	4.0E-04	120	76,975	66,996
	170	34.3	0.6	8.05E-02	6.1E-04	35.2	0.3	8.10E-02	5.0E-04	130	80,271	66,996
Hfq avg		33.6	0.5	4.45E-02		34.2	0.4	4.40E-02	4.0E-04	120	77,332	66,996
rpoS RNA	1.2	63.8	2.0	3.95E-02	1.3E-03	72.0	2.0	4.20E-02	1.2E-03	250	102,027	96,209
	2.3	63.8	1.2	7.80E-02	1.0E-03	71.5	1.4	8.20E-02	1.4E-03	250	105,402	96,209
	4.6	64.2	0.9	1.55E-01	2.0E-03	71.8	0.9	1.64E-01	1.8E-03	250	106,059	96,209
rpoS avg		66.0	1.6	8.97E-02	2.0E-03	76.7	2.7	9.60E-02	2.0E-03	300	106,642	96,209
1:1 Hfq:*rpoS*	2.3	58.0	1.0	1.38E-01	2.1E-03	61.1	0.7	1.40E-01	1.4E-03	250		163,205

SAXS data for Hfq protein, rpoS301 RNA and 1:1 Hfq:*rpoS* mixtures were acquired at room temperature over a momentum transfer range of $0.005 < q < 1.007$ Å^{-1} at APS 12-ID-B as described in Materials and Methods. The P(r) real space parameters were calculated for $q = 0$ to 0.5 using GNOM. The particle mass was estimated from the SAXS scattering curves by the method of Rambo and Tainer for $q = 0$ to 0.3 to reduce contributions of noise at high q.

(1) Rambo, R.P. & Tainer, J.A. 2013, "Accurate assessment of mass, models and resolution by small-angle scattering", *Nature,* vol. 496, no. 7446, pp. 477-481.

Structural models of Hfq

I first built an all-atom model of the full-length *E. coli* Hfq hexamer by appending disordered N-termini (resi 1-4) and C-termini (resi 66-102) to a crystallographic model of the stable Sm core (PDB ID: 4HT8) (Wang et al., 2011). I imposed a P6 symmetry constraint in the rigid body modeling program CORAL (Petoukhov et al., 2012) to model spatial arrangement of Hfq monomers and simulate the missing C-termini based on the structure library integrated in the program. The resulting model correctly rebuilt Hfq hexamer core as in crystal structure with a chi value = 0.36 (Fig. 4.5A, wheat). Interestingly, the model predicted that C-termini pointed towards the distal face, although the exact occupancy was not accurate, because every residue of C-termini was assigned as glycine and P6 symmetry assumed identical conformation for all six termini (Fig. 4.5A, pink).

To assign correct residue identities and allow the N- and C-termini conformation to vary, Dr. Joseph Curtis helped me to carry out a Monte Carlo simulation by using SASSIE program to simulate termini conformations that fit the experimental SAXS data (Materials and Methods) (Curtis et al., 2012). The trajectory, representing 24991 structures that distributed relatively symmetrically in space, covered the experimental Rg at the lowest X^2 values (Fig. 4.6A). In the best-fitting structure (Fig. 4.7A and B), N-termini and C-termini predominantly occupied the proximal- and distal-face, respectively. N-termini primarily covered the center pocket of proximal face, and occasionally flipped outwards to contact the rim (Fig. 4.7A, purple). By contrast, the C-termini sampled more diverse conformations in a large spatial envelope (Fig. 4.7A, pink). Intriguingly, the acidic terminal regions (resi 95-102) were predicted to be less disordered, and frequently interacted with positively charged residues on the rim or the proximal face (Fig. 4.7A), consistent with the fluorescent quenching data (Robinson et al., 2014). This distal orientation differs from the radial projection of the C-termini in previous *ab initio* models (Henderson et al., 2013) (Fig. 4.4C).

Structural models of *rpoS* leader

To model the tertiary structure of the free *rpoS* mRNA, I divided the *rpoS301* sequence into six fragments using my SHAPE-determined secondary structure as a guide (Fig. 4.5B): an upstream four-way-junction (red), a long helix connecting the $(AAN)_4$ and A_6 loops (orange), a short hairpin next to the $(AAN)_4$ motif (green), a linker region (blue), the inhibitory stem (purple), and the downstream four-way junction containing the U_5 motif (magenta). The single-stranded $(AAN)_4$ and A_6 motifs were assumed to be unstructured linkers (grey).

Three-dimensional models of *rpoS* mRNA secondary structure fragments (Fig. 4.5B) were generated using MC-Sym webserver (Parisien and Major, 2008). The predicted structures of *rpoS* fragments were consistent with the experimental SHAPE data (Chapter 3 Fig. 3.2A-B), with highly modified residues occurring in hairpin loops or kinks in the RNA backbone (Fig. 4.5B). The "hinge" region of the structure (light blue in Fig. 4.5B) is least well determined, but even the model for this region is in reasonable agreement with the data. The two 3-helix junctions (type C; (Lescoute and Westhof, 2006)) were also predicted using a knowledge-based method (Kim et al., 2014; Laing and Schlick, 2009; Laing et al., 2012). This method returned the same stacking geometry as in my model for the inhibitory stem-loop domain (pink in Fig. 4.5B), but a slightly different stacking geometry for the upstream 3-helix junction (red in Fig. 4.5B).

I then arranged the fragments in space by rigid-body modeling (SASREF; (Petoukhov and Svergun, 2005)) against the experimental SAXS data (Fig. 4.5C and Materials and Methods). This procedure constrained proximity between connecting nucleotides to be ~7-8 Å without restricting domain orientation. In the resulting model, the upstream and inhibitory domains again form an L connected by a flexible hinge at nt. 128-129 (Fig. 4.5C). Because these domains likely sample different orientations in solution, this hinge was used as a pivot point in a SASSIE Monte Carlo simulation, which generated an ensemble of 27,427 structures spanning the experimental R_g (Fig. 4.6B). The best-fit structures from this ensemble resembled the initial L-shaped model (Fig. 4.7 C).

Structural models of the *rpoS*•Hfq complex

I repeated this modeling procedure to visualize the structure of the *rpoS*•Hfq complex, using the scattering data from the 1:1 *rpoS*:Hfq_6 sample as an experimental constraint (Materials and Methods). I used a crystallographic structure of the Hfq core bound to rA_7 (Wang et al., 2011) to model the interaction between the $(AAN)_4$ motif and the distal face of Hfq. In addition, as my SHAPE data showed that the *rpoS* U_5 motif and A157 in the inhibitory stem both interact with the rim of Hfq, I constrained those residues to be within 7 Å of R16 in any Hfq monomer.

The resulting model (Fig. 4.7E and F) showed the *rpoS* mRNA wrapped around the Hfq hexamer, with the U_5 motif on the proximal side of the rim opposite the second

AAN triplet, and A157 at the rim on the other side of the ring. Strikingly, this orientation projected the inhibitory stem across the proximal face of Hfq, with the sRNA complementary strand toward Hfq and the ribosome binding site away from Hfq. This wrapped structure necessitates a slight unwinding of the inhibitory stem, consistent with the moderate increase in SHAPE modification of this region when Hfq binds. Hfq may induce additional RNA conformational changes that are not captured by my rigid body modeling procedure. Overall, the model explained how Hfq folds the *rpoS* mRNA into a more compact structure, and why interactions with both the AAN motif and the U_5 motif are needed for efficient sRNA entry.

To determine whether other conformations also fit the SAXS data, SASSIE was used to vary the orientation of the downstream RNA domain about the flexible hinge (nt. 128-129). Structures of the *rpoS*•Hfq complex that best represent the data ($X^2 < 1.5$; 917 structures) were symmetrically distributed about $R_g = 55$ Å (Fig. 4.7H), and collectively sampled a restricted wedge of space that could reflect an oscillatory path of the inhibitory stem in which the U_5 motif detaches and rebinds the Hfq lateral rim (Fig. 4.7G and Fig. 4.6D). This ensemble of “open” structures described the scattering data nearly as well as the initial “closed” structure (Fig. 4.7F). In all of these structures, nt A157 remained close to the Hfq rim, consistent with my SHAPE data showing that hyper-modification of this residue in the inhibitory stem depends on the $(AAN)_4$ motif binding rather than the U_5 motif. By contrast, the sRNA annealing site, ribosome binding site, and the U_5 motif moved away from Hfq in the more open structures.

Figure 4.5. Tertiary structures of rpoS301 fragments predicted by MC-Sym (Parisien and Major, 2008) and rigid body modeling (Petoukhov and Svergun, 2005). (*A*) CORAL rigid body model of free Hfq proteins. C-termini (pink) pointed towards the distal face of Hfq core (wheat). (*B*) Using the SHAPE-determined secondary structure as a guide (Fig. 3.2 Chapter 3), *rpoS301* was divided into six fragments. The tertiary structure of each fragment was predicted by MC-Sym (Parisien and Major, 2008): an upstream four-way-junction (red), a long helix connecting the $(AAN)_4$ motif and A_6 loops (orange), a short hairpin next to the $(AAN)_4$ motif (green), a linker region (blue), the inhibitory stem (purple), and the downstream four-way junction containing the U_5 motif (magenta). The single-stranded $(AAN)_4$ and A_6 motifs were assumed to be unstructured linkers, based on moderate SHAPE reactivity (grey). The predicted secondary structures were consistent with SHAPE modification (Fig. 3.2 Chapter 3). Highly modified nucleotides are shown as spheres on the ribbons. (*C*) SASREF rigid body model of free *rpoS* RNA. Left: Initial output of SASREF with disconnected fragments arranged to satisfy proximity constraints and the experimental SAXS scattering. Right: Final models were built after filling in missing fragments and connecting adjacent RNA modules. (*D*) CORAL rigid body model of the *rpoS*•Hfq complex, as in (*C*).

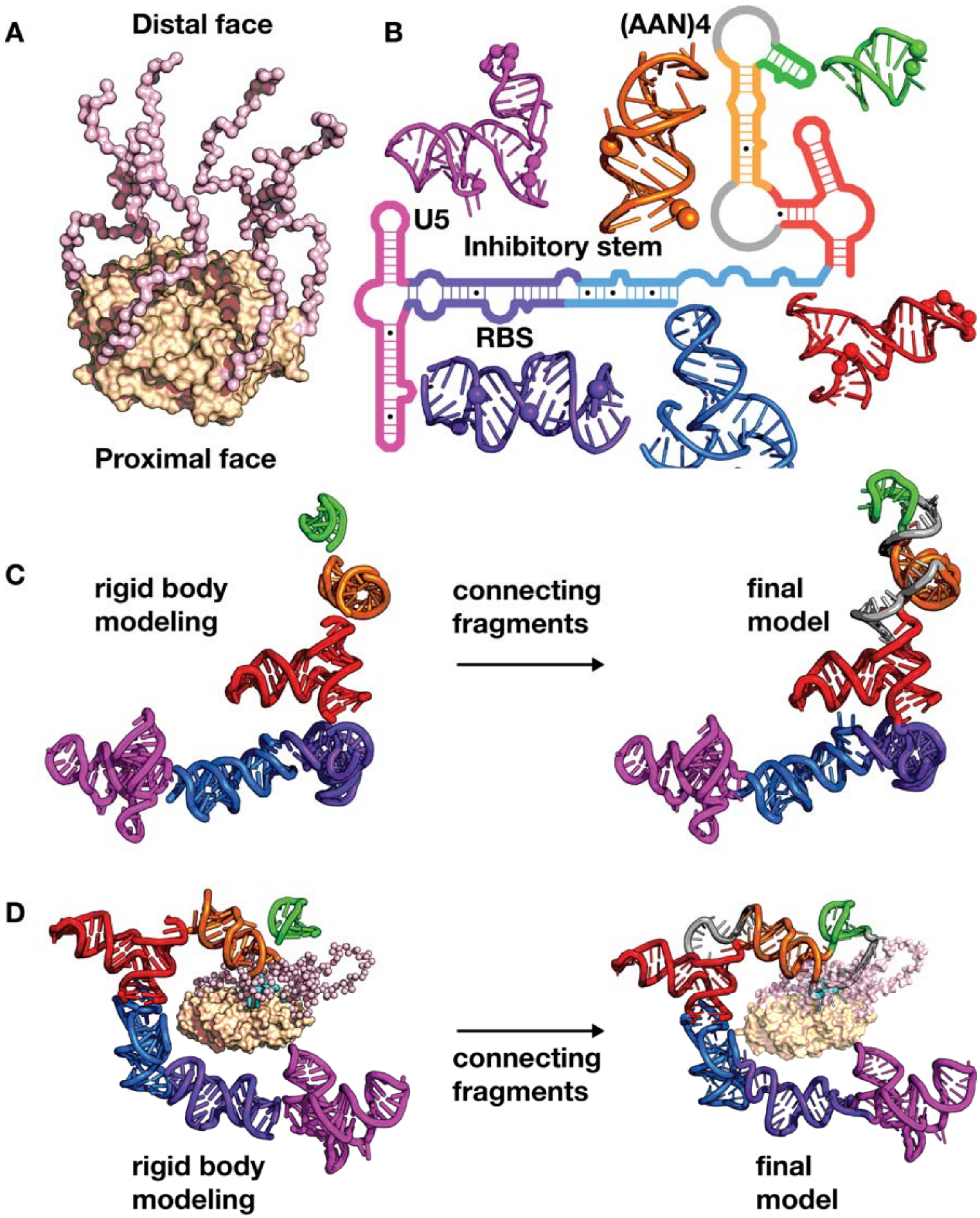

A
Distal face
Proximal face
B
(AAN)4
U5
Inhibitory stem
RBS
C
rigid body modeling
connecting fragments
final model
D
rigid body modeling
connecting fragments
final model

Figure 4.6. Monte-Carlo and Molecular Dynamics simulations of free Hfq and *rpoS* RNA (courtesy of Dr. Joseph E. Cutis and Dr. Sarah A. Woodson). (*A*-*C*) Comparison of experimental SAXS data with Monte Carlo simulations of (*A*) Hfq (24,991 structures), (*B*) free *rpoS* mRNA (27,427 structures), and (*C*) *rpoS*•Hfq complex (19,132 structures), performed with the program SASSIE (Curtis et al., 2012). In (*A*), the conformations of the intrinsically disordered N- and C-terminal residues were varied until the space around the hexamer core was fully sampled. (*B*,*C*) Residues 128-129 were used as a pivot point to produce an ensemble of structures spanning the experimental R_g for the free RNA (68.1 Å) and the complex (58 Å). The metric *Vr* (Rambo and Tainer, 2013) (blue in panel C) reaches a minimum at similar R_g values as chi-squared, but is less sensitive to variations in the structures. The "best-fit" structure in this ensemble is depicted as an example of the "open" conformation of the *rpoS*•Hfq complex (Fig. 4.7E). (*D*) Spatial distribution downstream *rpoS* mRNA domain in simulated Hfq•*rpoS* RNA structures (grey surface). Models were aligned to Hfq and superimposed. One conformation in the trajectory is shown as a ribbon; RNA, violet; Hfq, yellow. Top row, in the full trajectory, the downstream *rpoS* mRNA domain sampled the entire space around the Hfq Sm core. Bottom row, structures that best match the experimental SAXS data ($X^2 \leq 1.5$) are confined to a wedge of space around the proximal face of Hfq, as also shown in Fig. 4.7G. (*E*) A plot of the residual (calculated profile - experimental profile) for each of the 917 structures of the complex with $X^2 \leq 1.5$. While some positive serial correlation was observed (Durbin-Watson statistical test < 2) the magnitude of the residuals were on the order of the statistical error of the data.

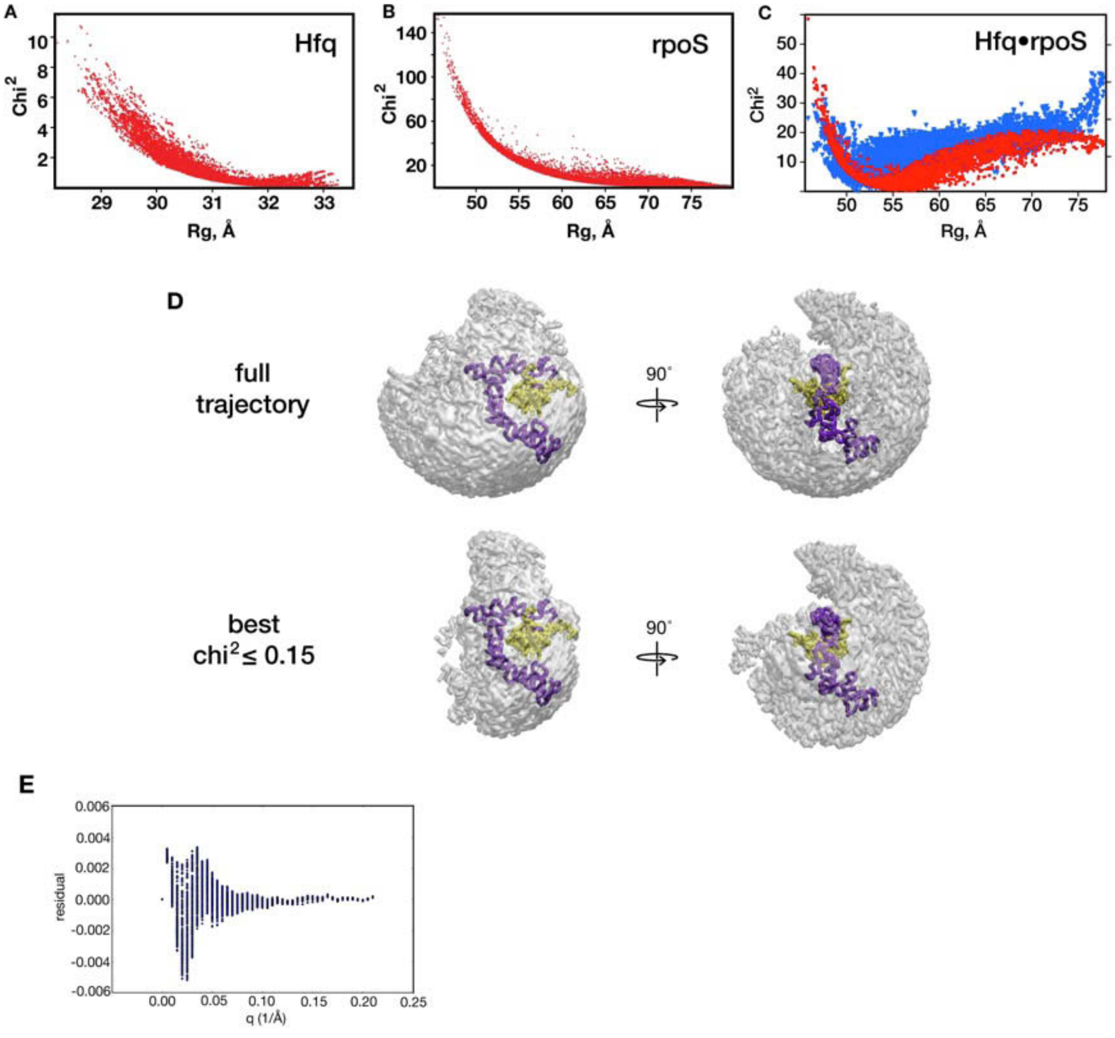
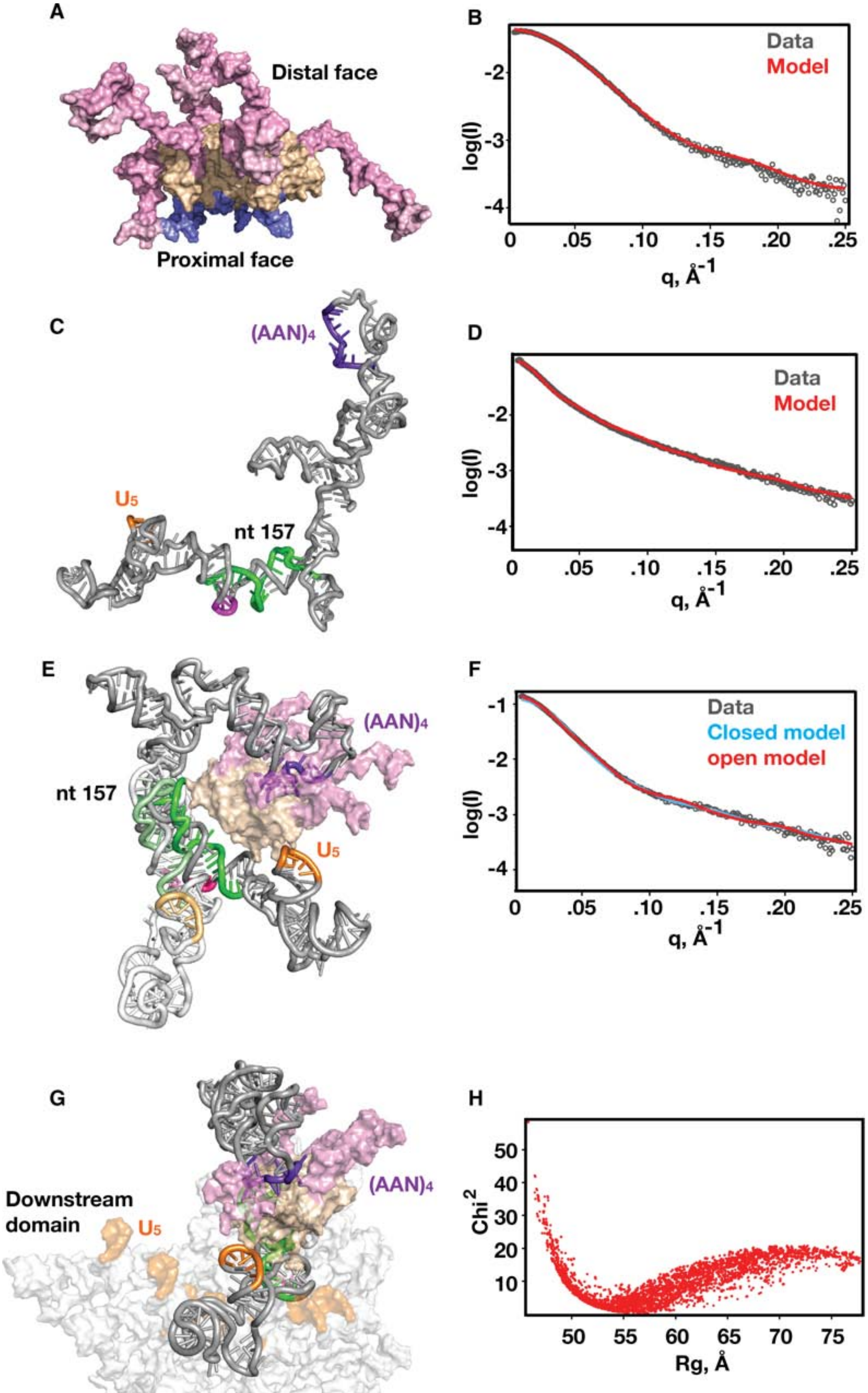

A
Hfq
Chi^2
Rg, Å
B
rpoS
Chi^2
Rg, Å
C
Hfq•rpoS
Chi^2
V(r)
Rg, Å
D
full trajectory
90°
best chi^2≤ 0.15
90°
E
residual
q (1/Å)

Figure 4.7. Model of the *rpoS* RNA•Hfq regulatory complex. All-atom models of (*A*,*B*) full-length Hfq, (*C*,*D*) *rpoS* RNA, (*E*,*F*) *rpoS*•Hfq complex. The U_5-rim contact was constrained in the closed model (SASREF; (Petoukhov and Svergun, 2005)); the open model is from the SASSIE (Curtis et al., 2012) trajectory. Hfq is rendered as a surface; Sm core (res 6-65), wheat; N-termini (res 1-5), slate; C-termini (res 66-102), pink. *rpoS* RNA ribbon in grey; $(AAN)_4$ motif, purple; U_5 motif, orange; sRNA binding site, green; Shine-Dalgarno site, violet. See Fig. 4.5 for details of the RNA model. (*B*, *D*, *F*) Scattering curves predicted by models (red or blue) compared with experimental scattering (gray) for (*B*) full-length Hfq ($X^2 = 0.27$), (*D*) *rpoS* RNA ($X^2 = 0.58$), (*F*) *rpoS*•Hfq complex in "closed" or "open" conformation ($X^2 = 0.43$). (*G*) Open structures from SASSIE. Best fitting 917 models from the trajectory (*H*) were clustered (UCSF Chimera (Yang et al., 2012)), and each cluster represented by a semi-transparent surface to illustrate the wedge of conformations that describe the scattering curve.


A
Distal face
Proximal face
B
Data
Model
log(I)
-2
-3
-4
0
.05
.10
.15
.20
.25
q, Å^{-1}
C
(AAN)$_4$
U$_5$
nt 157
D
Data
Model
log(I)
-2
-3
-4
0
.05
.10
.15
.20
.25
q, Å^{-1}
E
(AAN)$_4$
nt 157
U$_5$
F
Data
Closed model
open model
log(I)
-1
-2
-3
-4
0
.05
.10
.15
.20
.25
q, Å^{-1}
G
(AAN)$_4$
Downstream domain
U$_5$
H
Chi^2
50
40
30
20
10
50
55
60
65
70
75
Rg, Å

Discussion

My SAXS data and all-atom models collectively show that Hfq folds the *rpoS* mRNA leader into a compact tertiary structure in solution. This folded structure positions the inhibitory stem of the *rpoS* leader over the proximal face of Hfq where sRNAs are known to bind. This result explains the detailed mechanisms of *rpoS* regulation by sRNAs and Hfq, such as how Hfq brings together the complementary regions of the mRNA and sRNA near the arginine patches along the rim, and why sequences upstream and downstream of the sRNA target site are important. Moreover, in agreement with previous evidence showing that Hfq partially opens the secondary structure of the inhibitory stem to enhance sRNA annealing and ribosome binding (Hammerle et al., 2013; Soper et al., 2011), remodeling of the *rpoS* mRNA at tertiary structure level also requires interactions with both $(AAN)_4$ and U_5 motifs.

As the SAXS data do not provide information about local structure, my model cannot capture the details of the RNA-Hfq interactions. Moreover, the model does not account for local perturbations to the RNA structure because RNA fragments were treated as rigid entities in the modeling procedure. Nevertheless, the overall arrangement of the *rpoS* mRNA leader with respect to Hfq in my model is well supported by experimental data. First, the dramatic change in the scattering function provides direct physical evidence for compaction of the RNA by Hfq. Second, the remarkable change in RNA backbone modification (SHAPE) in response to Hfq binding and various Hfq and *rpoS* mutations is consistent with specific Hfq interactions, rather than non-specific effects of the protein on the RNA structure (Chapter3). Third, mutational studies showed that the position and orientation of the $(AAN)_4$ and U_5 sequences are important for Hfq-mediated sRNA regulation, suggesting they bind Hfq simultaneously (Chapter3; (Peng et al., 2014)). Finally, an unbiased search of structural models indicated that only a subset of RNA conformations recapitulate the SAXS data (Fig. 4.7G and H and Fig. 4.6D).

Rigid body modeling show that Hfq folds the *rpoS* leader into a compact, “closed” conformation by simultaneously recognizing an upstream $(AAN)_4$ motif and downstream U_5 motif flanking the sRNA target site. In this closed mode, the inhibitory stem is partially melted, and the 5’ end of the target site interacts with the Hfq rim where the arginine patch was proposed to promote base pairing with a complementary sRNA (Panja et al., 2013). The SHAPE data show that Hfq disengages from the downstream U_5 motif after a sRNA base pairs with the inhibitory stem, while remaining bound to the $(AAN)_4$ motif (Chapter 3). The potential to form more open structures explains how the *rpoS* leader can flex to allow Hfq to cycle off the DsrA-*rpoS* duplex, exposing the ribosome binding site.

The potential for opening and closing the *rpoS*•Hfq complex is clearly captured in the SASSIE models. The closed *rpoS*•Hfq model obtained by constraining the U_5 motif to interact with the Hfq rim was reasonably consistent with the SAXS data. However, the Monte Carlo simulations showed that more open structures fit the scattering data equally well, even assuming a small fraction of free RNA. In fact, calculated R_g value of the open conformation is closer to the experimental value than that of the closed model, further supporting the conclusion that an ensemble of *rpoS*•Hfq structures may fluctuate between open and closed conformations in solution.

In all the modeling procedures, *rpoS* and Hfq were assumed to form 1:1 complex. However, SAXS scattering curves for higher Hfq:*rpoS* molar ratios also indicate a folded structure, suggesting the existence of higher order conformations. In fact, my data do not exclude models in which the open *rpoS* leader binds a second Hfq hexamer. Although including a second Hfq hexamer in rigid body modeling against SAXS scattering of higher order Hfq:*rpoS* complex did not generate reasonable structure, it is likely due to the lack of biochemical data to constrain proximity of the second Hfq to the other components.

While AAN sequences are known to recruit Hfq via its distal face (Link et al., 2009; Mikulecky et al., 2004; Soper and Woodson, 2008), here I find that the U_5 motif in *rpoS* also contributes to sRNA annealing by interacting with the Hfq rim. This distorts the mRNA structure, making it more accessible to sRNAs (Soper et al., 2011). Multi-lateral Hfq interactions may be widespread among bacterial sRNA-mRNA pairs and important for regulation. The *fhlA* mRNA leader was proposed to contact both distal and proximal faces of Hfq based on competitive binding experiments (Salim and Feig, 2010). Hfq inhibits translation of *cirA* by binding to an upstream (AAN) motif and two U-rich patches close to the Shine-Dalgarno sequence (Salvail et al., 2013), raising the possibility that Hfq also folds the *cirA* mRNA for translational control. My results show that Hfq forms a specific, folded *rpoS* mRNP that spring-loads the regulatory helix for sRNA entry.

Hfq-induced structural change not only applies to the mRNA leaders but also occurs to sRNAs. A newly established Hfq-RNA interaction model proposed that Hfq directly recognizes the sRNA 3’ U-tail on the proximal surface and interacts with the U-rich patches within the sRNA body on its lateral surface (Sauer, 2013). This multi-surface interaction induced sRNA structural change, revealed by the sRNA shape change in the low-resolution molecular envelope upon complex formation (Henderson et al., 2013). Moreover, this structural rearrangement protected the sRNAs from RNase E degradation, highlighting a mechanism to control sRNA level and potentially sRNA pairing with target mRNAs (Henderson et al., 2013). Further studies are necessary to investigate how Hfq interacts with various sRNAs, which will be discussed in Chapter 5.

Materials and methods

SAXS sample preparation

Hfq was purified as previously described (Peng et al., 2014). Concentrated protein was dialyzed twice against SAXS buffer (annealing buffer plus 2 mM $MgCl_2$) at a final concentration of 1.90 mg/ml (170 µM monomer). *rpoS301* RNA was purified by native 6% polyacrylamide gel in 1X THEM2 , eluted from the gel overnight at 4 °C in SAXS buffer, concentrated by ultrafiltration (Amicon Ultra-15 centrifugal filter unit, 50KD), and washed five times with fresh SAXS buffer (Fang et al., 2013). The final concentration was 0.43 mg/ml (4.6 µM). Samples were shipped on ice and stored at 4 °C before use. Remaining SAXS buffer from the sample preparation was used for diluting samples and measuring background scattering. Small angle X-ray scattering data were collected at the Advanced Photon Source 12-ID-B, over the range $0.005 < q < 1.007$ Å^{-1} as described previously (Fang et al., 2013). Guinier fits and real space inversions were done using Primus and GNOM from the ATSAS software package (Konarev et al., 2003). Other plots were generated using the ScÅtter software package (bl1231.als.lbl.gov/scatter).

A small amount of RNA dimer (11% of dynamic light scattering; <1% number density) was detected for the RNA stock (1.9 mg/ml) that was stored at -80 ˚C (Fig. 4.2). Nevertheless, SAXS collected on fresh (never frozen) samples with difference concentrations showed constant R_g and ratios of scattering intensity, indicating a lack of inter-particle interactions (Fig. 4.1). Estimates of the molecular mass by the method of Rambo and Tainer (Ramos et al., 2013) were within 10% of the nominal value (Table

4.1). The disordered C-terminus of Hfq may result in higher than expected estimates by this method (Watson and Curtis, 2014).

SAXS data collection

Small angle X-ray scattering data were collected at room temperature at the Advanced Photon Source 12-ID-B, over the range $0.005 < q < 1.007$ Å^{-1} after background subtraction. Data collected at three different sample concentrations showed the expected increase in I(0) and constant R_g and ratios of scattering intensity, indicating a lack of inter-particle interactions (Fig. 4.1). Parameters of the fits and estimates of the particle mass are listed in Table 4.1.

MC-Sym

Three-dimensional models of *rpoS* mRNA secondary structure fragments (Fig. 4.5) were generated using MC-Sym webserver (Parisien and Major, 2008). The RNA sequence and secondary structure based on SHAPE experiments were used as the input with default settings; two-stranded fragments were first connected with a GAAA tetra-loop. The output structures were ranked using the MC-Sym webserver tools. The top five predicted structures for each fragment did not differ substantially. The highest scoring structure for each fragment was selected for rigid body modeling, after removal of GAAA tetra-loops and energy minimization with UCSF Chimera (Yang et al., 2012).

Rigid body modeling of *rpoS* and *rpoS*•Hfq complex

The tertiary structure of *rpoS* RNA was modeled by orienting the RNA fragments by rigid-body modeling (SASREF) (Petoukhov and Svergun, 2005) using the SAXS experimental data as a constraint ($X^2 = 0.59$). The RNA connectivity was enforced by setting the distance between adjacent phosphorous atoms to ≤ 7 Å (resi 40-41, 57-58, and 127-128). To allow more flexibility in the fitting procedure, the distance between connecting phosphorus atoms was constrained to ≤ 8 Å in only one of the two strands in the inhibitory stem (resi 162-163 or 272-273 for connecting fragments 4 and 5, 184-185 or 248-249 for connecting fragments 5 and 6). Because I could not model single-stranded regions of the upstream domain (resi 73-86, 102-116) with MC-Sym, the missing sequences were initially built assuming an A-form conformation, and all the fragments were manually connected into a continuous strand in UCSF Chimera (Yang et al., 2012). Structures resulting from repeated calculations were similar, with two structural domains oriented at approximately 90°. The predicted structures were not sensitive to changes in distance constraints.

CORAL was used to model the full *rpoS*•Hfq complex against the SAXS data for the 1:1 RNA:Hfq sample (Petoukhov et al., 2012). An initial model of Hfq core bound to the AAN_4 motif was obtained by importing the coordinates of six protomers of *E. coli* $Hfq_{5\text{-}65}$ in complex with A_7 RNA (PDB ID: 4HT8) (Wang et al., 2011). To this model, I added the missing C-termini of Hfq (resi 66-102), and the *rpoS* fragments used for SASREF modeling of free *rpoS* RNA with the same contact restraints. The A7 RNA bound to Hfq was joined to the rest of the RNA by setting the distance between *rpoS301* P 88 and the 3' end of A7 ≤ 7 Å. Two additional constraints were introduced based on the

SHAPE footprinting data: *rpoS301* P 195 (U_5 motif) ≤ 12 Å from the α-carbon of R16 in any Hfq monomer and *rpoS301* P157 (inhibitory stem) ≤ 15 Å from R16 α-carbon in any Hfq monomer. Finally, the missing nucleotides (nt 73-80, 102-116) were built and connected manually in UCSF Chimera.

Monte Carlo simulations of Hfq, *rpoS*, and Hfq•*rpoS* complexes by SASSIE

Monte Carlo simulations (SASSIE; (Curtis et al., 2012)) were used to identify conformations of free Hfq, free *rpoS* mRNA, and the *rpoS*•Hfq complex consistent with the scattering data for each sample. The coordinates of the Hfq core (alone or with rA bound to the distal face (Wang et al., 2011)) were fixed during the simulations, while the N- and C-terminal residues (resi 1-5 and resi 66-102) were allowed to vary. The initial full-length Hfq structure was energy minimized using CHARMM (Brooks et al., 2009), before it was used as the input for the Monte Carlo simulation. During the Monte Carlo simulation, Crysol 2.7 (Svergun et al., 1995) was used to calculate scattering profiles of simulated structures after renaming atoms to C, H, N, O, P, S to avoid reading errors in Crysol. The averaged SAXS profile of Hfq was interpolated (43 points, $\Delta q = 0.005$ from 0.005 to 0.21) and was used to evaluate the theoretical scattering profiles. Models generated by SASSIE that best fit the experimental SAXS data were minimized using CHARMM.

The free *rpoS* structure generated by rigid body modeling was modeled using SASSIE as described above. However, the RNA was allowed to pivot around the flexible hinge connecting the upstream and downstream domains (resi 128-129). The Monte Carlo dihedral sampling of RNA backbone configurations were carried out using

CHARMM 36 force-field parameters (Foloppe and MacKerell, 2000) for a, b, g, d, e, h angles using the same energetic sampling as described previously (Curtis et al., 2012).

The RNA coordinates of the *rpoS*•Hfq complex generated from rigid body modeling and the full-length Hfq structure generated from SASSIE were merged into a starting structure for Monte Carlo simulations of the complex. CHARMM was used to minimize the energy in three steps by first restraining all RNA atoms, then the RNA carbon and phosphate atoms, and then allowing all atoms to move. A SASSIE Monte Carlo simulation was carried out as above, using *rpoS* resi 128-129 as pivot point and allowing Hfq N- and C-termini to vary. The SAXS profile of the1:1 *rpoS*•Hfq sample was used to evaluate the theoretical scattering profiles, using reduced chi-squared as a measure of statistical goodness-of-fit. An alternative measure of error V(r) (Ramos et al., 2013) reached a minimum around a similar range of R_g values (Fig. 4.6). The best-fit structures were energy minimized using CHARMM. Final models were compared to the scattering data using FoXS (Schneidman-Duhovny et al., 2013).

Structural models

Three-dimensional models of *rpoS* mRNA secondary structure fragments (Fig. 4.5) were generated using MC-Sym webserver (Parisien and Major, 2008), and oriented in three-dimensions with SASREF (Petoukhov and Svergun, 2005) using the RNA chain connectivity and the SAXS experimental data as constraints. CORAL was used to model the full *rpoS*•Hfq complex against the SAXS data for the 1:1 RNA:Hfq sample (Petoukhov and Svergun, 2005). In the complex, *rpoS* P 195 (U_5 motif) and P 157 (inhibitory stem) were constrained to ≤ 12 or 15 Å, respectively, from R16 Cα in any Hfq monomer. Monte Carlo simulations were performed using the program SASSIE (Curtis et al., 2012) to identify conformations of free Hfq, free *rpoS* mRNA, and the *rpoS*•Hfq complex consistent with the scattering data for each sample. The coordinates of the Hfq core were fixed during the simulations, while the N- and C-termini (aa 1-5 and aa 66-102) were allowed to move. The RNA was allowed to pivot between nt 128-129. While the residuals between the best 917 models and the experimental data for the 1:1 *rpoS*•Hfq complex showed some positive serial correlation (Durbin-Watson < 2), the magnitudes of the residual were on the order of the statistical error of the data (Fig. 4.6E). Detailed methods were described above.

Chapter 5 Distinct Hfq binding modes for sRNAs with A-rich and U-rich motifs

Introduction

Bacterial regulatory small RNAs (sRNAs) play important roles in essential cellular processes (Waters and Storz, 2009). sRNAs are usually ~100 nt long transcripts that act in trans to regulate translation of mRNAs transcribed from other parts of the genome (Gottesman and Storz, 2011). Most sRNAs base pair with the sequence close to the ribosome binding site in the target mRNAs, leading to translation inhibition and destabilization of the mRNAs(Storz et al., 2011). In several cases, sRNAs up-regulate mRNA translation by resolving an inhibitory secondary structure that occludes ribosome binding (Storz et al., 2011). In many bacteria such as *E.coli*, the interaction between the sRNA and the mRNA is mediated by the RNA chaperone Hfq (Vogel and Luisi, 2011).

The current model for Hfq-sRNA interaction posits that the proximal surface of Hfq directly binds to the sRNA 3’ U-tail and its lateral surface interacts with the sRNA body (Sauer, 2013). However, this U-rich binding model was only biochemically demonstrated for the model sRNA RybB, which does not contain A-rich repeats (Sauer et al., 2012). In fact, there is evidence that RprA sRNA interacts with both the proximal and distal surfaces of Hfq, while OxyS associates with the proximal, distal, and lateral surfaces(Henderson et al., 2013; Updegrove and Wartell, 2011). Moreover, single mutations on the RNA binding surfaces of Hfq differentially affected the intracellular levels of various sRNAs and translational regulation of target mRNAs, suggesting a variety of sRNA-Hfq binding modes (Zhang et al., 2013). However, it remains largely unknown that how Hfq binds to various sRNAs containing A- and U-rich sequences and how Hfq facilitates annealing between various RNA pairs.

Here, I studied four sRNAs that contain either A- or U-rich motifs to illustrate how Hfq differentially recognizes the sRNAs and mediates strand annealing with the mRNA targets. *In vitro* Hfq binding assays showed that ChiX and RprA sRNAs were less affected than other sRNAs by a mutation on the proximal face of Hfq. They also did not depend on a 3' U-tail for tight binding, whereas DsrA and McaS require a U-tail to form a stable complex with Hfq. RNase footprinting revealed structural changes oin sRNAsbound to WT Hfq and the Hfq mutants, suggesting that each sRNA binding mode requires different coordination of multiple Hfq surfaces. Finally, co-transcription annealing experiments suggested that Hfq's action mode depends on the complexity of the RNA secondary structure and the complementarity of the sRNA-mRNA pairs.

Results

Hfq binds to sRNAs in distinct modes

To investigate how Hfq binds sRNAs containing either A-rich or U-rich motifs, I used ChiX, DsrA, McaS, and RprA sRNAs that were differentially stabilized by various Hfq alleles *in vivo* (Zhang et al., 2013). All four sRNAs contain a U-rich tail, which is a common feature of Rho-independent transcripts. ChiX and RprA also include an A-rich motif within the sRNA body whereas DsrA and McaS contain some short U-rich patches (Fig. 5.1).

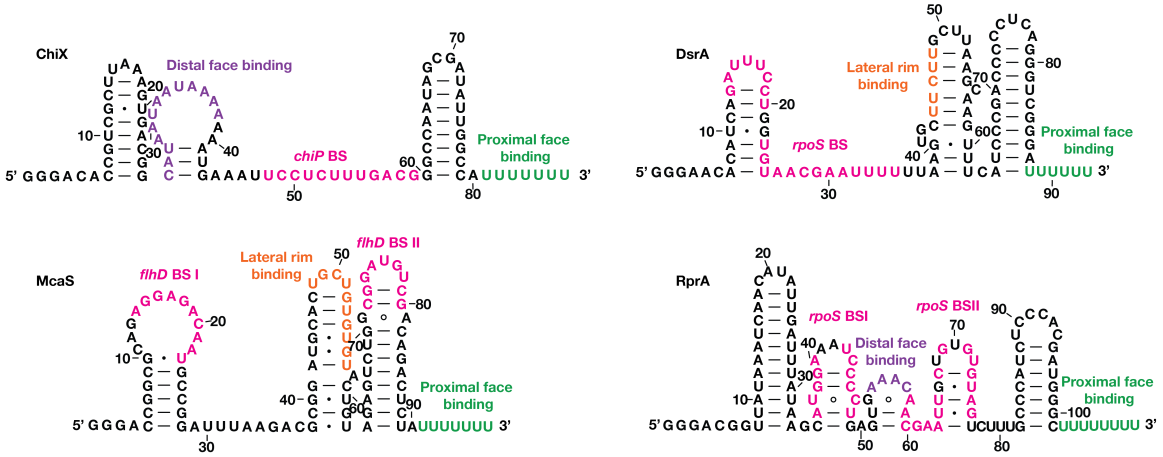


Figure 5.1. Secondary structures of ChiX, DsrA, McaS, and RprA sRNAs. The Structures were generated by Mfold webserver (Zuker, 2003). A-rich motifs binding to Hfq distal face were colored in purple; internal U-rich motifs binding to Hfq lateral surface were colored in orange; 3'-U tails binding to Hfq proximal face were colored in green; mRNA annealing sites were colored in magenta.

I first measured Hfq binding constants of the sRNAs by native polyacrylamide gel electrophoresis (Fig. 5.2). ^{32}P-labeled sRNAs were titrated with wt Hfq and Hfq variants containing the rim mutation R16A, the distal face mutation Y25D, the proximal face mutation K56A, and the C-terminal truncation Hfq65. All four sRNAs showed binding curves that were best described by a tight binding site and a nonspecific binding site (Fig. 5.2 and Table 5.1). ChiX showed the highest affinity for wt Hfq with dissociation constant $K_d = 0.17$ μM (Fig. 5.2A). Moreover, this strong Hfq binding retained in all Hfq variants with K_d values ranging from 0.12 – 0.21 μM, suggesting that either ChiX binds to Hfq at multiple sites simultaneously or ChiX has a very strong single Hfq binding site (Fig. 5.2 A and Table 5.1). If ChiX-Hfq binding relies on the A-rich motif as suspected, then truncating U-tail should not affect the binding constant much. In fact, ChiX-ΔU showed the same binding affinity to wt Hfq ($K_d = 0.15$ μM) as full-length ChiX, confirming that ChiX binding does not depend on the U-tail for its affinity to Hfq.

As expected, other sRNAs showed a drastically different Hfq binding profile (Fig. 5.2 B-D and Table 5.1). All the other three sRNAs bind to wt Hfq with a slightly higher dissociation constant ($K_d = 0.20\text{-}0.25$ μM). The K56A mutation reduced the binding affinity by 2-5 fold, suggesting that the Hfq proximal face interaction is required for sRNA binding. Interestingly, DsrA and McaS sRNAs bound to Hfq very poorly without the U-tail ($K_d = 1.31$ and 9.13 μM, respectively), consistent with the model that Hfq interaction mainly comes from the U-tail binding at the proximal surface. By contrast, RprA binding was only slightly reduced by the U-tail truncation ($K_d = 0.28$ μM), suggesting that RprA may contain additional binding sites within the sRNA body.

The Y25D mutant binds all the sRNAs surprisingly well with K_d ranges from 0.12–0.15 μM (Fig. 5.2 purple curves and Table 5.1). It is conceivable that this single mutation may not be sufficient to disrupt the distal-face binding site for A-rich motifs. Alternatively, Y25D mutation changes the electrostatic landscape on the Hfq surface for nonspecific binding, which plausibly explained higher order complex formation in the native gel (Fig. 5.2).

Another interesting observation comes from Hfq65 binding (Fig. 5.2). Deletion of the C-termini strongly reduced the homogeneity of ChiX and McaS complexes in the gel, although the binding constant did not change significantly. This effect was not observed for DsrA and RprA, suggesting that Hfq C-termini may stabilize complex conformation for some sRNAs.

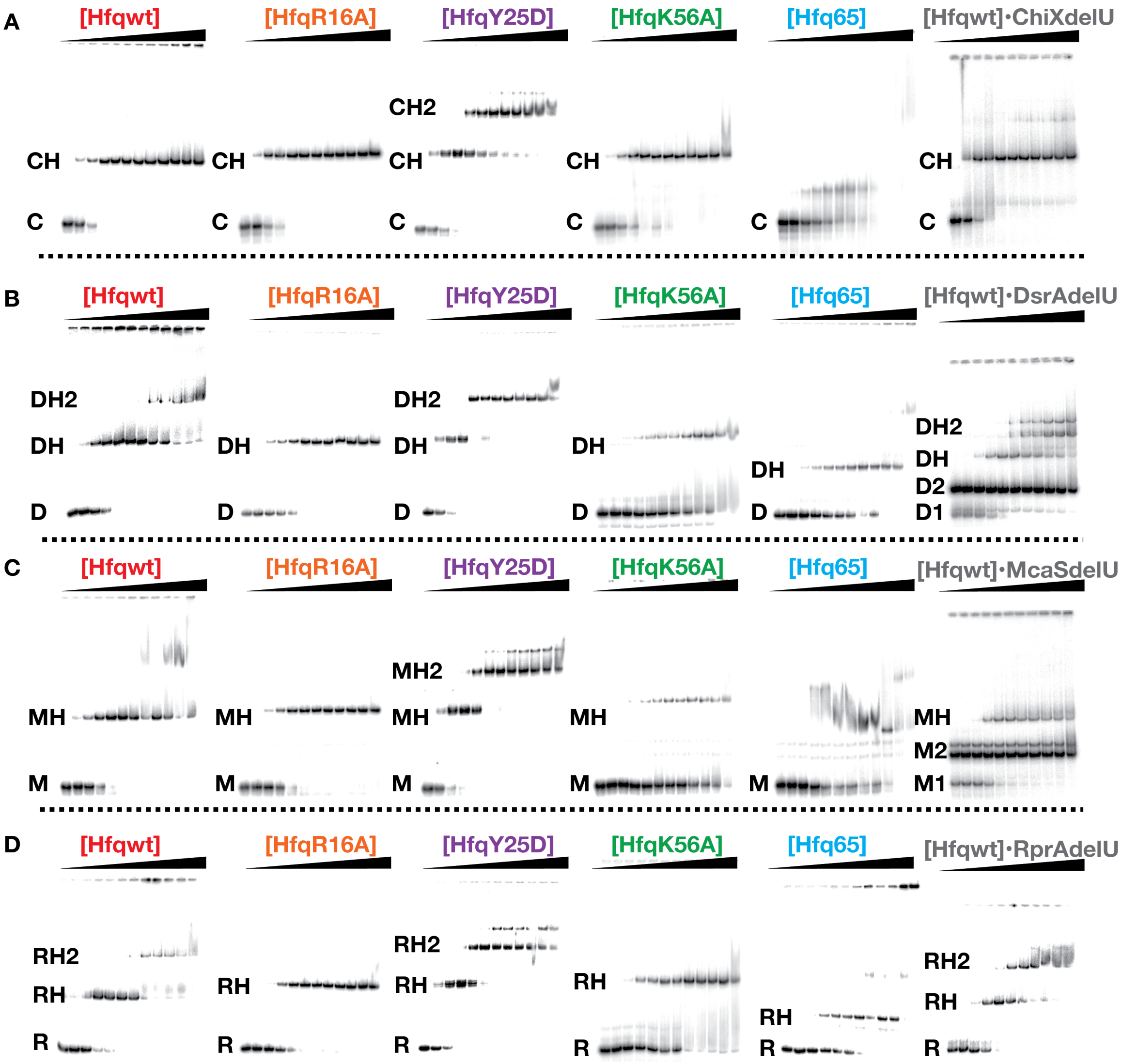


Figure 5.2. Measuring Hfq binding constants for sRNAs. Hfq variants titrations of uniformly labeled (*A*) ChiX, (*B*) DsrA, (*C*) McaS, and (*D*) RprA sRNAs at 25 ºC. Free sRNA (lower band) binds one or two Hfq multimers (first and second band shifts). U-tail truncation of sRNAs were titrated with wt Hfq. Hfq concentration ranges from 0 –2 µM monomers (Materials and Methods).

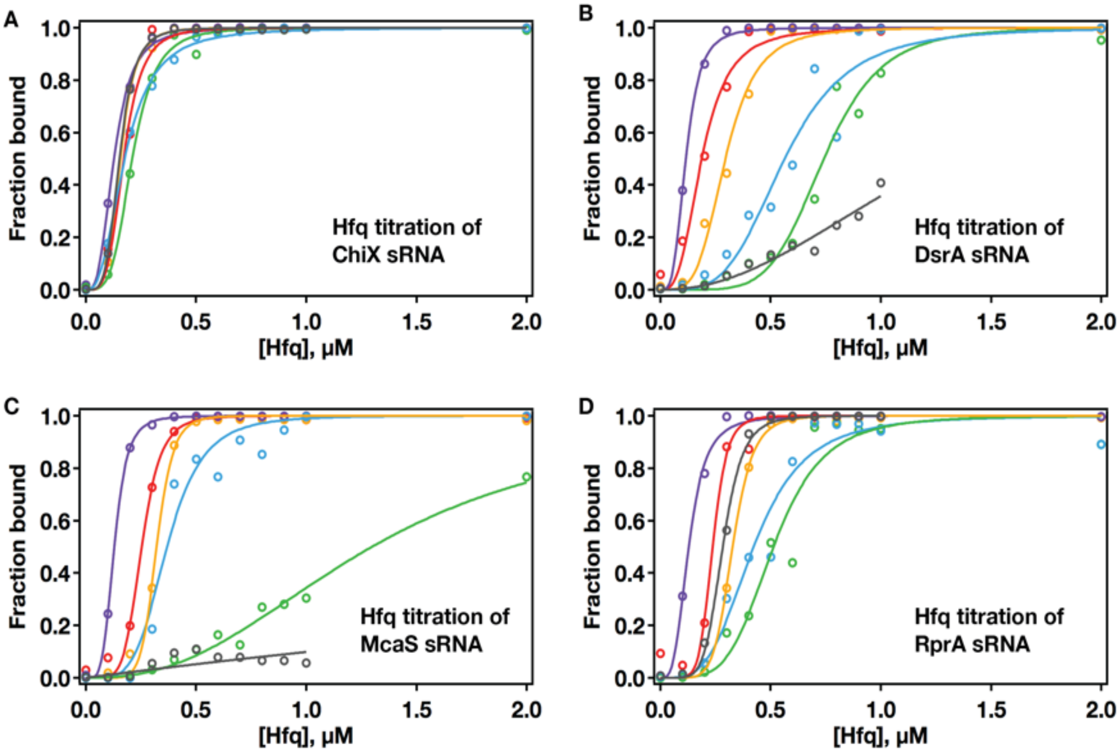


Figure 5.3. Fraction of bound sRNA as a function of [Hfq]. For comparison purpose, the total fraction bound was plot by combining the fractions of sRNA bound to one or two Hfq multimers (Fig. 5.2); this total fraction was fit with a partition function assuming 1 binding site (K_d). For more accurate K_d estimation for sRNAs that bind two Hfq multimers, fractions of RNA•Hfq and RNA•Hfq2 were quantified separately and fit with a partition function assuming two binding sites (K_{d1} and K_{d2}) (Materials and Methods); combined K_d did not differ significantly from K_{d1} (Table 5.1). Full length (*A*) ChiX, (*B*) DsrA, (*C*) McaS, and (*D*) RprA sRNAs were titrated with wt Hfq (Red), R16A (orange), Y25D (purple), K56A (green), and Hfq65 (blue). 3'-U tail truncation of sRNAs were titrated with wt Hfq (grey).

Table 5.1. Summary of sRNA•Hfq binding constants

		Hfq binding constant K_d (μM)			
		ChiX	DsrA	McaS	RprA
WT Hfq	K_d	0.25 ± 0.06	0.19 ± 0.01	0.25 ± 0.004	0.24 ± 0.007
	K_{d1}	-	0.20 ± 0.02	-	0.24 ± 0.03
	K_{d2}	-	0.86 ± 0.06	-	0.53 ± 0.01
R16A		0.16 ± 0.002	0.29 ± 0.01	0.32 ± 0.004	0.33 ± 0.002
Y25D	K_d	0.13 ± 0.002	0.12 ± 0.01	0.13 ± 0.001	0.13 ± 0.003
	K_{d1}	0.12 ± 0.007	0.15 ± 0.03	0.12 ± 0.008	0.12 ± 0.01
	K_{d2}	0.44 ± 0.01	0.35 ± 0.05	0.41 ± 0.02	0.41 ± 0.03
K56A		0.21 ± 0.005	0.74 ± 0.02	1.30 ± 0.04	0.51 ± 0.03
Hfq65		0.18 ± 0.003	0.56 ± 0.03	0.37 ± 0.02	0.42 ± 0.02
sRNA-ΔU		0.15 ± 0.001	1.31 ± 0.10	9.13 ± 12.2	0.28 ± 0.003

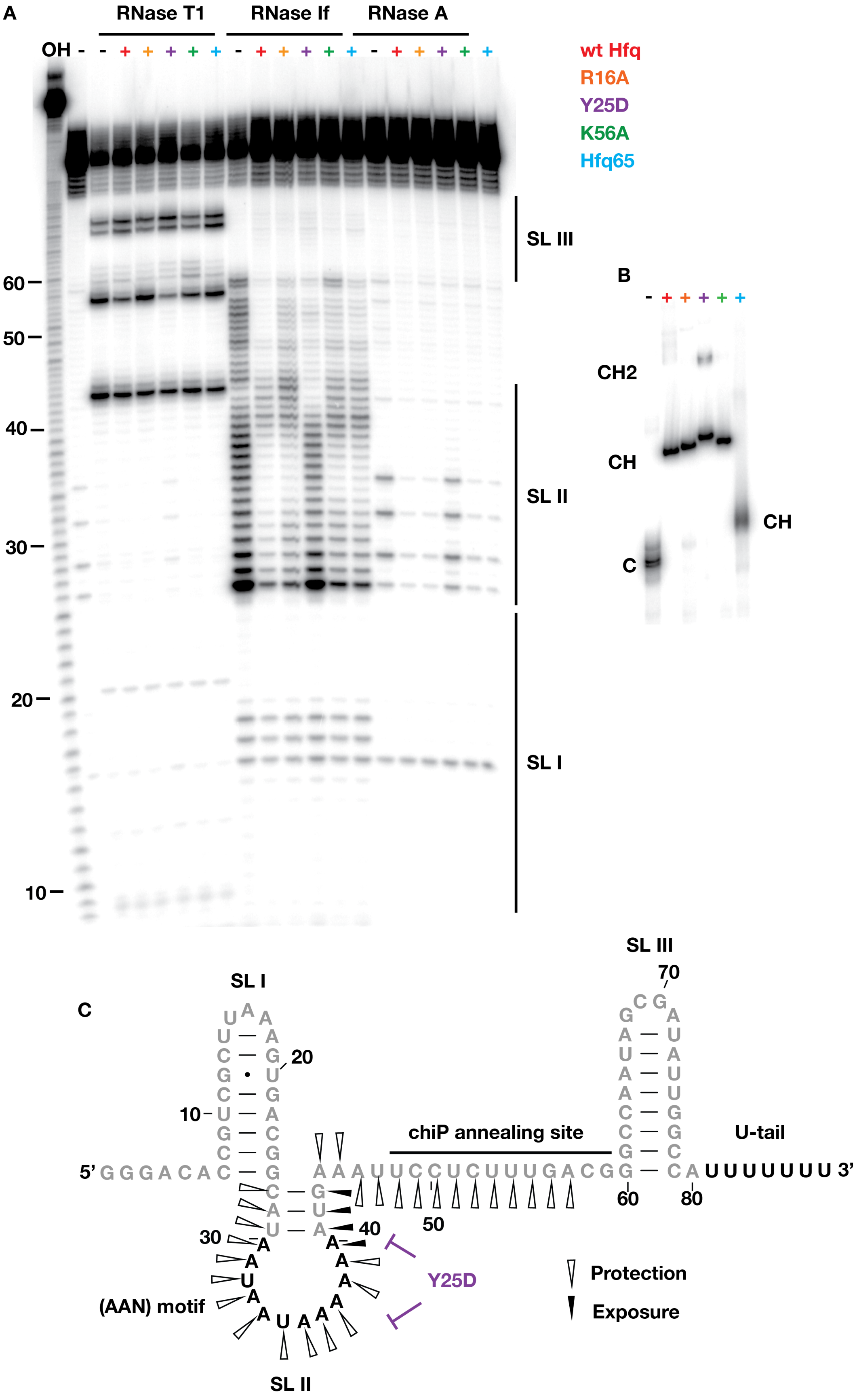
Hfq binding constants were measured by native polyacrylamide gel electrophoresis (Fig. 5.2 and 5.3). The equations and calculations were described in Material and Methods. Errors for ChiX•wt Hfq was calculated as the standard deviation of K_d values from three replicate measurements; errors for the other reactions were calculated as the fitting errors by IGOR Pro.

Hfq binding changes sRNA structure

Since Hfq mutations affect binding affinity to various sRNAs differently, I further investigated where and how Hfq binds to the sRNAs by using a combination of RNases to probe the sRNA structures (Fig. 5.4-5.7). RNase T1 cleaves 3’ of single stranded G nucleotides; RNase A cleaves 3’ of single stranded U and C nucleotides; and RNase If cleaves all RNA dinucleotide bonds with a preference for single-stranded RNA (Ehresmann et al., 1987). Hfq binding may protect the binding site from RNase digestion or induce structural changes in other regions that reduce or enhance cleavage of the sRNA.

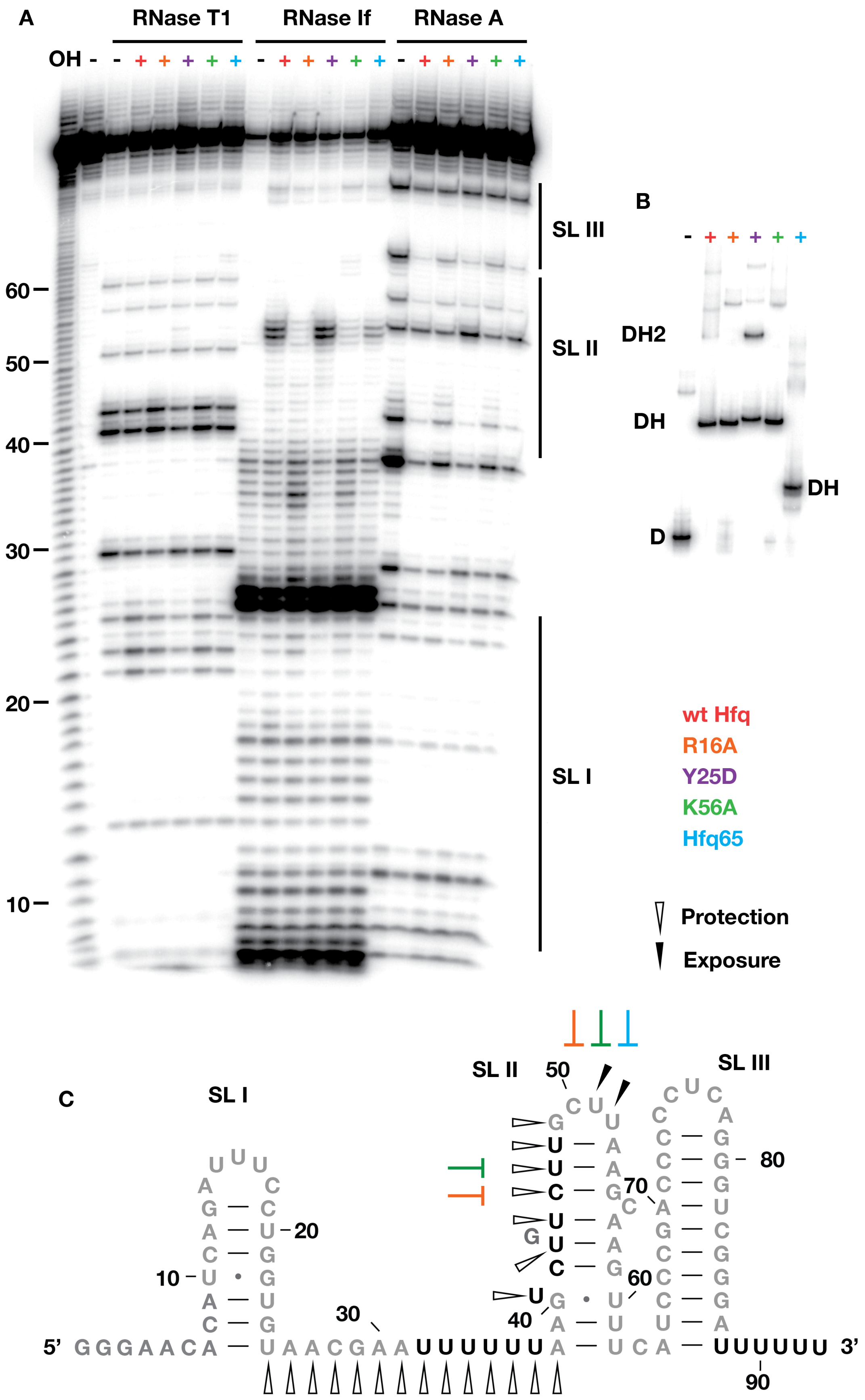
Hfq was proposed to recognize an A-rich motif in stem-loop (SL) II of ChiX sRNA (Fig. 5.4A and C). This stem-loop was highly protected by wt Hfq and other Hfq variants except Y25D (Fig. 5.4A), suggesting that the SL II A-rich motif directly binds to Hfq distal face. Strikingly, Hfq Y25D still bound ChiX tightly and formed a higher order complex as shown in the native gel (Fig. 5.4B). It is likely that ChiX binds tightly to other Hfq surfaces in those complexes. In addition, while all Hfq variants completely bound ChiX sRNA and protected the single-stranded region between SL II and SLIII, the protection is slightly weaker for K56A and R16A proteins (Fig. 5.4A and C). One possibility is that Hfq may wrap ChiX sRNA around its core, which has been observed for OxyS and RprA sRNAs (Henderson et al., 2013).

Figure 5.4. RNase footprinting of ChiX sRNA. (*A*) 5'-labeled ChiX sRNA alone or in complex with Hfq wt (red), R16A (orange), Y25D (purple), K56A (green), and Hfq65 (blue) were partially digested by RNase T1, If, and A followed by sequencing polyacrylamide gel electrophoresis. Sequencing lanes were generated as previously described (Lease and Woodson, 2004). Note that the OH-ladder sample of ChiX sRNA was switched with RprA sRNA by mistake. (*B*) Undigested reactions were resolved on a native polyacrylamide gel to verify complex formation. (*C*) Summary of RNase footprinting profile of ChiX sRNA in complex with Hfq variants as compared to ChiX alone. (AAN) motif and 3'-U tail are highlighted in black. mRNA annealing site is indicated. Reduced or increased of RNase digestion at specific nucleotide is indicated by open or solid wedge. Y25D disrupted RNase digestion pattern of SL II as indicated.


A
RNase T1
RNase If
RNase A
OH - - + + + + + - + + + + + - + + + + +
wt Hfq
R16A
Y25D
K56A
Hfq65
60
50
40
30
20
10
SL III
SL II
SL I
B
- + + + + +
CH2
CH
CH
C
C
SL I
SL III
70
20
10
chiP annealing site
U-tail
5' G G G A C A C
C A U U U U U U U 3'
60
80
30
40
50
Y25D
(AAN) motif
SL II
Protection
Exposure

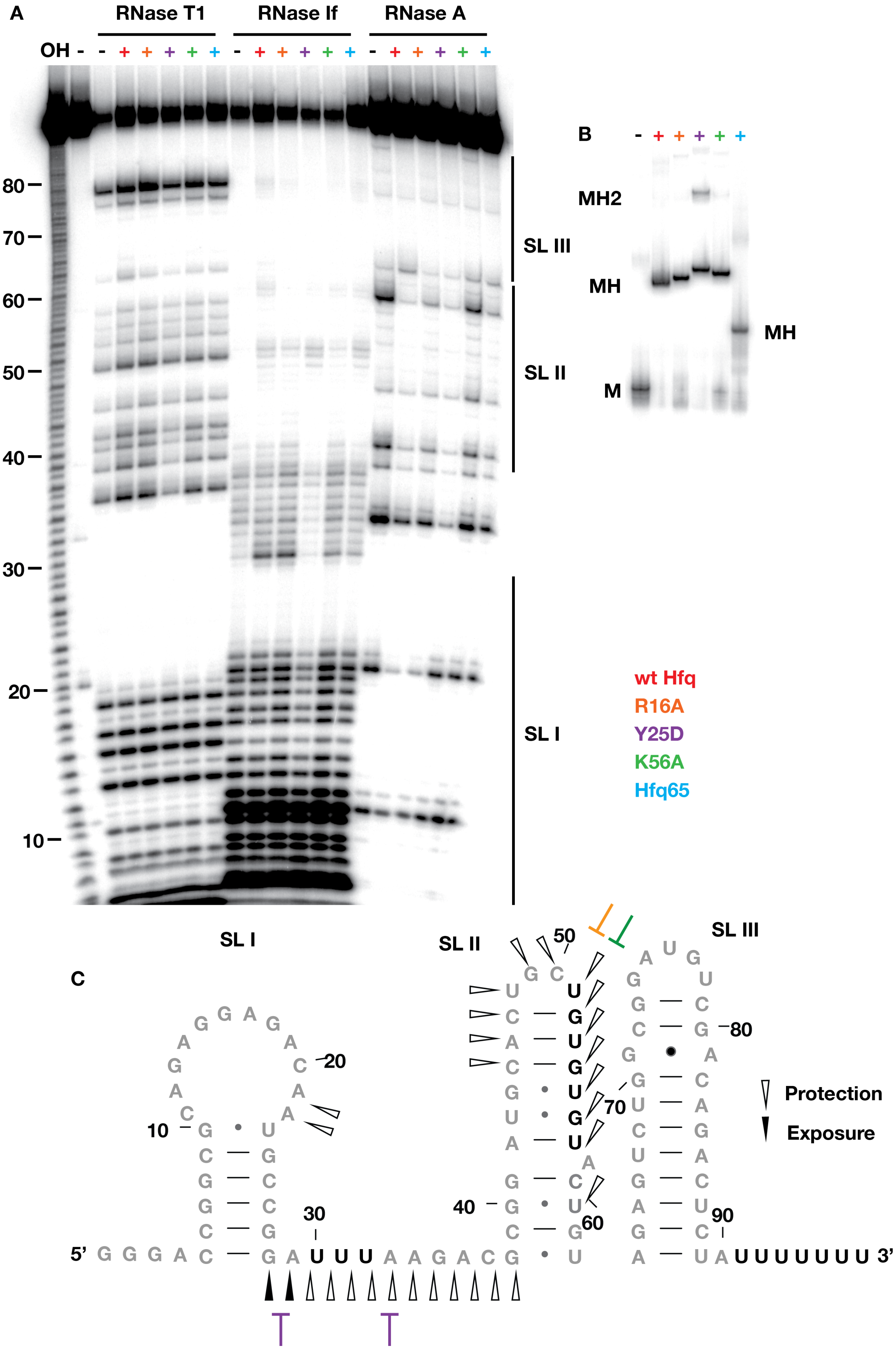
This is different from DsrA sRNA, which has been known to bind Hfq by the internal U-patch and the 3'-U tail (Lease and Woodson, 2004; Updegrove and Wartell, 2011). In my RNase footprinting data (Fig. 5.5A and C), the 5' half of SL II was slightly protected from RNase A by wt Hfq and Hfq Y25D but less so by Hfq R16A, K56A, and Hfq65, suggesting that the Hfq rim and proximal surface directly binds the U-rich patch. The internal binding also destabilized the stem-loop structure and induced RNase If digestion at the loop region of SL II. Another major change in RNase digestion pattern occurred at the single-stranded region between SL I and SL II, which was protected by all Hfq variants from RNase A digestion. These observations were consistent with previous RNase footprinting results for DsrA sRNA that contains a shorter U-tail (Lease and Woodson, 2004).

Figure 5.5. RNase footprinting of DsrA sRNA. (*A*) 5'-labeled DsrA sRNA alone or in complex with Hfq wt (red), R16A (orange), Y25D (purple), K56A (green), and Hfq65 (blue) were partially digested by RNase T1, If, and A followed by sequencing polyacrylamide gel electrophoresis. (*B*) Undigested reactions were resolved on a native polyacrylamide gel to verify complex formation. (*C*) Summary of RNase footprinting profile of DsrA sRNA in complex with Hfq variants as compared to DsrA alone. Internal U-rich motifs and 3'-U tail are highlighted in black. Reduced or increased of RNase digestion at specific nucleotide is indicated by open or solid wedge. Hfq mutations that altered the RNase digestion patterns were indicated.


A
RNase T1
RNase If
RNase A
OH - - + + + + + - + + + + + - + + + + +
60
50
40
30
20
10
SL III
SL II
SL I
B
- + + + + +
DH2
DH
DH
D
wt Hfq
R16A
Y25D
K56A
Hfq65
Protection
Exposure
C
SL I
SL II
SL III
5' GGGAACA
UUUUUU 3'
10
20
30
40
50
60
70
80
90

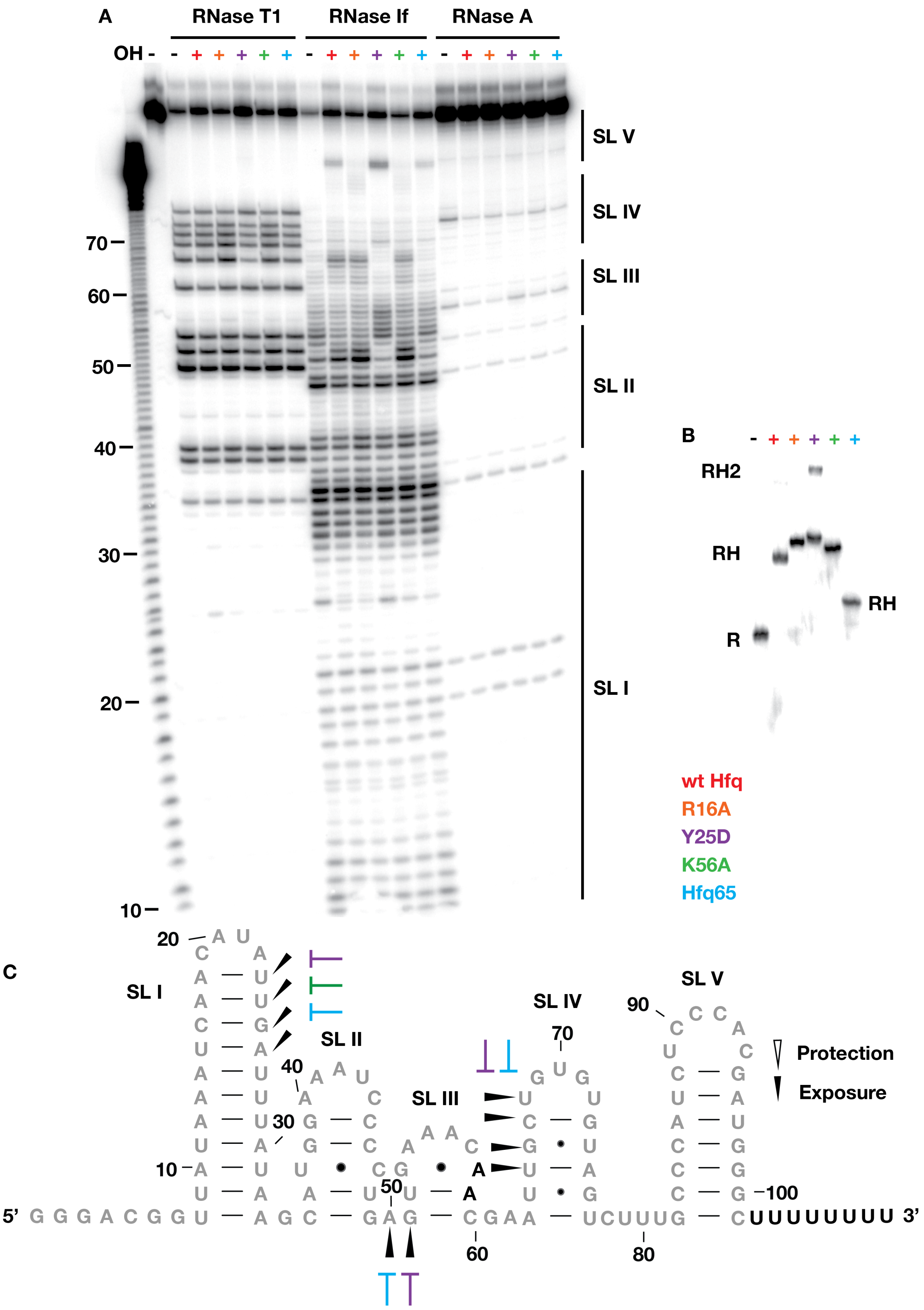
Similar to DsrA, McaS SL II also interacted with Hfq's lateral surface and proximal face (Fig. 5.6A and C). This U- and C-rich patch was protected from RNase A digestion whereas the joint between SL II and SL III was more exposed by Hfq wt and Y25D, suggesting that Hfq binding to SL II partially unwinds the adjacent secondary structure. Interestingly, the single-stranded region between SL I and SL II was significantly protected from RNase A and RNase If digestion by Hfq Y25D, raising the possibility that altered Hfq binding mode leads to nonspecific binding or changed RNA conformation, consistent with the higher order complex formation for Y25D (Fig. 5.6B). Finally, Hfq also induced subtle structural change at the loop region and bottom of SL I, which was impaired by Hfq distal and proximal face mutations.

Figure 5.6. RNase footprinting of McaS sRNA. (*A*) 5'-labeled McaS sRNA alone or in complex with Hfq wt (red), R16A (orange), Y25D (purple), K56A (green), and Hfq65 (blue) were partially digested by RNase T1, If, and A followed by sequencing polyacrylamide gel electrophoresis. (*B*) Undigested reactions were resolved on a native polyacrylamide gel to verify complex formation. (*C*) Summary of RNase footprinting profile of McaS sRNA in complex with Hfq variants as compared to McaS alone. Internal U-rich motifs and 3'-U tail are highlighted in black. Reduced or increased of RNase digestion at specific nucleotide is indicated by open or solid wedge. Hfq mutations that altered the RNase digestion patterns were indicated.


A
RNase T1
RNase If
RNase A
OH
80
70
60
50
40
30
20
10
SL III
SL II
SL I
B
MH2
MH
MH
M
wt Hfq
R16A
Y25D
K56A
Hfq65
C
SL I
SL II
SL III
5'
3'
10
20
30
40
50
60
70
80
90
Protection
Exposure

Finally, RprA underwent significant structural change upon Hfq binding at SL III, IV, and V (Fig. 5.7A and C). Based paired regions in those short stems were more exposed to RNase If digestion upon binding with Hfq wt, R16A, and K56A, whereas Hfq Y25D and Hfq65 increased exposure to RNase If at the loop region of SL III. Together with the high-order complex formation for Y25D in the native gel (Fig. 5.7B), it is possible that different Hfq binding modes induced differential structural changes. Surprisingly, I did not observe an internal region significantly protected by Hfq as expected for an sRNA that does not depend on U-tail binding (Fig. 5.2 and 5.3) and binds to Hfq distal face(Updegrove et al., 2008). This could result from dynamic Hfq binding to multiple short patches in the sRNA body or the limitation of the RNase footprinting assay.

Figure 5.7. RNase footprinting of RprA sRNA. (*A*) 5'-labeled RprA sRNA alone or in complex with Hfq wt (red), R16A (orange), Y25D (purple), K56A (green), and Hfq65 (blue) were partially digested by RNase T1, If, and A followed by sequencing polyacrylamide gel electrophoresis. Note that the OH-ladder sample of ChiX sRNA was switched with RprA sRNA by mistake. (*B*) Undigested reactions were resolved on a native polyacrylamide gel to verify complex formation. (*C*) Summary of RNase footprinting profile of RprA sRNA in complex with Hfq variants as compared to DsrA alone. 3'-U tail is highlighted in black. Reduced or increased of RNase digestion at specific nucleotide is indicated by open or solid wedge. Hfq mutations that altered the RNase digestion patterns were indicated.


A
RNase T1
RNase If
RNase A
OH
SL V
SL IV
SL III
SL II
SL I
70
60
50
40
30
20
10
B
RH2
RH
RH
R
wt Hfq
R16A
Y25D
K56A
Hfq65
C
SL I
SL II
SL III
SL IV
SL V
Protection
Exposure
5'
3'
20
40
30
10
50
60
70
80
90
100

Hfq facilitates cotranscription annealing of sRNA and mRNA

So far I determined three different Hfq binding modes: ChiX internal A-rich motif binds to the Hfq distal surface, DsrA and McaS internal U-rich and 3' U-tails bind to the Hfq lateral rim and proximal surface, and RprA undergoes a significant structural change that is disrupted by single mutation on various Hfq surfaces. To investigate how the Hfq binding mode determines Hfq's chaperone activity of mRNA•sRNA annealing, I measured the rate of co-transcriptional annealing of various sRNA and mRNA pairs. *In vitro* transcription reactions of mRNAs were assembled in the presence of preformed ^{32}P-labeled sRNAs• Hfq complex. Aliquots of the reactions were resolved by native polyacrylamide gel electrophoresis at various times after the start of transcription to measure the kinetics of sRNA binding to the mRNA (Fig. 5.8).

In all the sRNA•mRNA pairs I tested, wt Hfq increased the annealing rate by at least 6-fold (Fig. 5.9 and Table 5.2). Moreover, single mutations reduced Hfq's annealing activity by at least 50%. This reduction in annealing rate was more substantial than the moderate effect of these Hfq mutations on sRNA binding affinity (Table 5.1 and 5.2), suggesting that the Hfq surfaces are actively involved in chaperone activity and not simply RNA binding sites.

More interestingly, the relative annealing activity of the Hfq variants tested showed distinct profiles that corresponded to the extent of complementarity between each sRNA-mRNA pair (Fig. 5.9 and Table 5.2). In the first group (Fig. 5.9A and B), the annealing activity was reduced by the Y25D and K56A mutations on the distal and proximal surfaces of Hfq to the level of no Hfq, and less affected by the rim R16A mutation, which retained moderate annealing activity. The RNA pairs in this group contained a single complementary region of ~20 bp. By contrast, the second group of RNA pairs contained multiple discontinuous regions of complementarity which were each ~10 bp. The annealing activity of these RNA pairs depended strongly on interactions with the Hfq proximal face, because the K56A mutant was the most deleterious, while the R16A and Y25D mutants showed moderate annealing activity (Fig. 5.9C-E). Finally, ChiX and chiP annealing site is a single-stranded 12 nt region. Annealing of these two RNAs strongly depended on the lateral surface because R16A did not enhance annealing compared to no Hfq (Fig. 5.9F).

This classification does not correlate with a specific RNA. For instance, McaS sRNA annealing to *csgD* and *flhD* mRNAs fell into group I and group II profiles, respectively (Fig. 5.9B and C). Similarly, *rpoS* mRNA annealing to DsrA and RprA sRNAs also followed different profiles (Fig. 5.9A and D). The reason is that Hfq's annealing activity in this assay does not depend on its binding affinity to a specific RNA but rather on how it structures the RNA for base pairing. For instance, Hfq:Y25D did not effectively recruit McaS sRNA to the *flhD* transcript although this protein bound McaS tightly (Fig. 5.8C). One plausible explanation is that Hfq:Y25D adopted an altered binding mode that did not restructure McaS for effective annealing (Fig. 5.6). This impaired restructuring ability had differential effect on mRNA targets with different annealing regions (Fig. 5.9B and C). Additional high-resolution structure mapping is required to establish an accurate correlation between the structural change caused by Hfq and its annealing activity.

Figure 5.8. Measuring cotranscriptional annealing rates for sRNA-mRNA pairs. Transcription reactions of mRNA were carried out in the presence of ^{32}P-labeled sRNA and Hfq variants to measure annealing rates for (*A*) ChiX-chiP, (*B*) DsrA-rpoS, (*C*) McaS-flhD, and (*D*) RprA-rpoS. All the sRNAs initially bind T7-RNA polymerase (Cpol, Dpol, Mpol, and Rpol) because the bands appeared after T7-RNA polymerase was added to all the reaction without Hfq present (first row, no Hfq reaction). The presence of Hfq prevented polymerase binding in some cases as indicated in the gels. DsrA formed multiple bands after being released from T7 RNA polymerase, suggesting that dissociation from T7 RNAP may cause DsrA sRNA to misfold. sRNAs bound with Hfq (CH, DH, MH, and RH) gradually annealed with mRNA target to form binary and ternary complexes. Some Hfq mutations disrupt ternary complex formation as indicated in the gels.

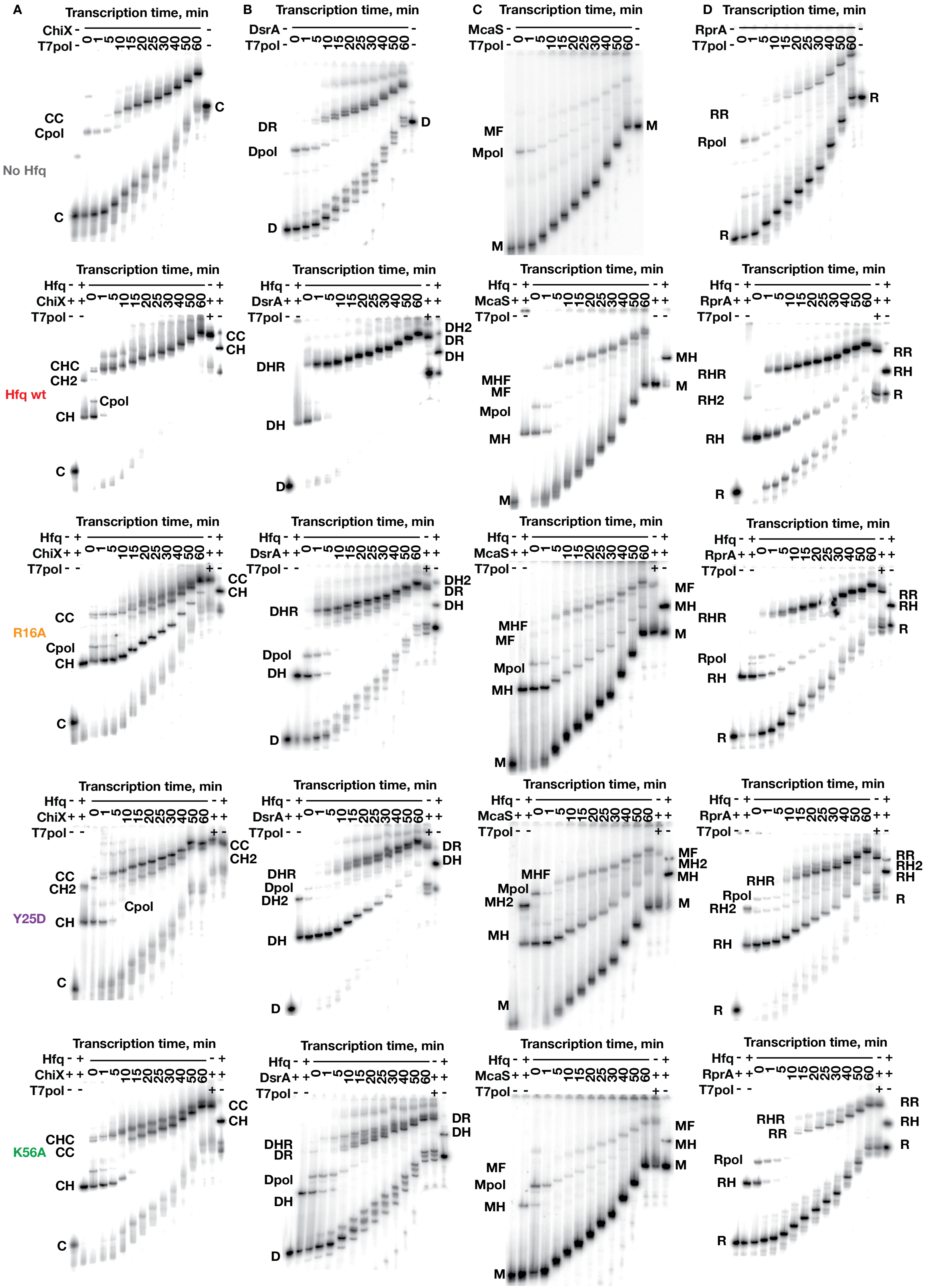
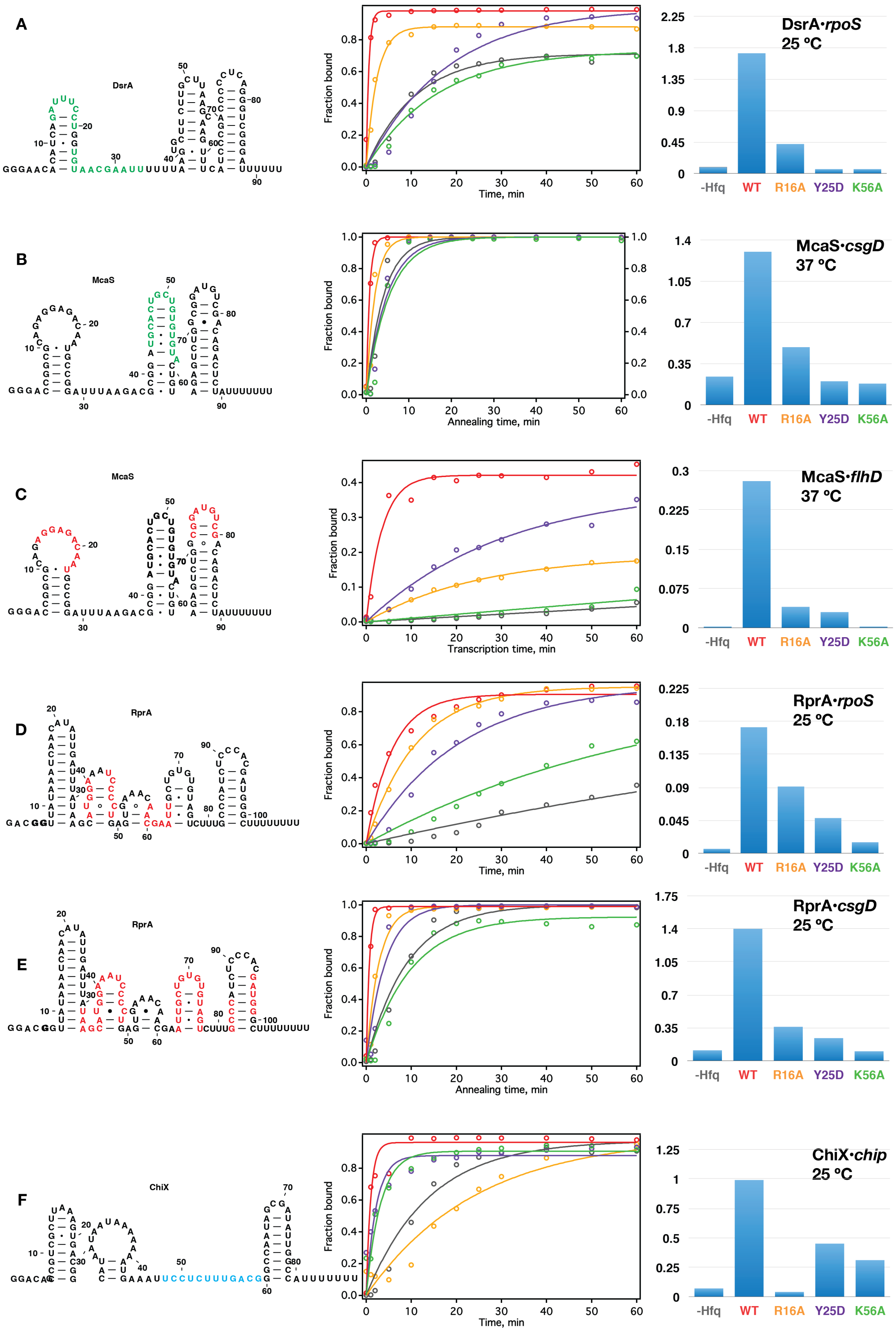

A
B
C
D
Transcription time, min
ChiX
DsrA
McaS
RprA
T7pol
Hfq
0 1 5 10 15 20 25 30 40 50 60
No Hfq
Hfq wt
R16A
Y25D
K56A
C
CC
Cpol
CH
CH2
CHC
D
DR
Dpol
DH
DH2
DHR
M
MF
Mpol
MH
MH2
MHF
R
RR
Rpol
RH
RH2
RHR

Figure 5.9. Cotranscriptional annealing rates of sRNA and mRNA pairs. Fraction bound was quantified from the data shown in Fig. 5.8 and additional experiments not shown. Left panels, annealing kinetics followed a monophasic exponential rise, not accounting for a ~1 min lag in mRNA synthesis; middle panel, comparison of annealing rates in presence of Hfq variants; right panels, secondary structures of sRNAs showing mRNA annealing site. (*A*) DsrA-rpoS and (*B*) McaS-csgD have a continuous region of base complementarity (green letters) within a stem-loop in the sRNA (Majdalani et al., 1998; Majdalani et al., 2001; Sledjeski et al., 2001; Thomason et al., 2012). (*C*) McaS-flhD, (*D*) RprA-rpoS, and (*E*) RprA-csgD contain multiple discontinuous complementary regions (red letters) (De Lay and Gottesman, 2012; Holmqvist and Vogel, 2013; Jorgensen et al., 2013; Majdalani et al., 2002; Mika et al., 2012; Teitel, 2000; Urban and Vogel, 2007). (*F*) ChiX-chiP base pair at single-stranded region in both RNAs (cyan letters) (Figueroa-Bossi et al., 2009; Overgaard et al., 2009). Fractions bound were fit with monophasic reaction (Materials and Methods). Annealing rate constants are summarized in Table 5.2.



A
DsrA
GGGAACA — UAACGAAUUUUUUA — UCA — UUUUUU
Fraction bound
Time, min
-Hfq
WT
R16A
Y25D
K56A
DsrA·*rpoS*
25 ºC
2.25
1.8
1.35
0.9
0.45
0
B
McaS
GGGAC — GAUUUAAGACG • U A — UAUUUUUUU
Annealing time, min
McaS·*csgD*
37 ºC
1.4
1.05
0.7
0.35
0
C
McaS
Transcription time, min
McaS·*flhD*
37 ºC
0.3
0.225
0.15
0.075
0
D
RprA
GACGGU — AGC — GAG — CGAA — UCUUUG — CUUUUUUUU
Time, min
RprA·*rpoS*
25 ºC
0.225
0.18
0.135
0.09
0.045
0
E
RprA
Annealing time, min
RprA·*csgD*
25 ºC
1.75
1.4
1.05
0.7
0.35
0
F
ChiX
GGACAG — G C — GAAAUUCCUCUUUGACGG — CAUUUUUUU
Time, min
ChiX·*chip*
25 ºC
1.25
1
0.75
0.5
0.25
0
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60

Table 5.2. Summary of sRNA•mRNA cotranscription annealing rat

	Annealing rate (min^{-1})				
	No Hfq	Hfq wt	R16A	Y25D	K56A
DsrA•rpoS	0.10 ± 0.01 (1)	1.72 ± 0.32	0.42 ± 0.03	0.06 ± 0.01 (1)	0.07 ± 0.01 (1)
McaS•csgD	0.26 ± 0.06 (1)	1.30 ± 0.09	0.48 ± 0.10	0.22 ± 0.05 (1)	0.20 ± 0.05 (1)
McaS•flhD	0.0019 ± 0.0001 (1)	0.28 ± 0.05	0.043 ± 0.004 (1)	0.036 ± 0.006 (1)	0.0028 ± 0.003 (1)
RprA•rpoS	0.006 ± 0.009	0.07 ± 0.006	0.09 ± 0.005	0.05 ± 0.008	0.014 ± 0.01
RprA•csgD	0.11 ± 0.02 (1)	1.44 ± 0.09	0.36 ± 0.06	0.23 ± 0.07 (1)	0.11 ± 0.02 (1)
ChiX•chiP	0.08 ± 0.01 (1)	0.96 ± 0.19 (2)	0.03 ± 0.008 (1)	0.20 ± 0.02 (2)	0.20 ± 0.03 (2)
DsrA•rpΔ2	0.08 ± 0.01 (1)	0.44 ± 0.02	-	0.07 ± 0.01 (1)	0.05 ± 0.01 (1)
RprA•rpoSΔ2	0.01 ± 0.0006 (1)	0.096 ± 0.008	-	0.05 ± 0.004 (1)	0.009 ± 0.001 (1)

sRNA •mRNA cotranscription annealing rate were measured by native polyacrylamide gel electrophoresis, as shown in Fig. 5.8-5.10. The equations and calculations were described in Material and Methods. Annealing reactions were fit with single exponential equations (Chapter 2 Materials and Methods eq. 2.3); some reactions were adjusted with either (1) a lag time term (eq. 5.1) or (2) an amplitude term (eq. 5.2). Errors were calculated as the fitting errors by KaleidaGraph.

Nevertheless, this cotranscriptional annealing profile reflected how Hfq acted on the sRNA rather than the mRNA to facilitate annealing. When mutating the A-rich motifs in the rpoS mRNA (*rpoSΔ2*), Hfq did not have any effect on post-transcriptional annealing rate for *rpoSΔ2* and DsrA RNAs (Peng et al., 2014; Soper and Woodson, 2008), consistent with the model that the A-rich motif is required for Hfq's to facilitate annealing. However, when measured during transcription, Hfq wt increased the annealing rate of the same RNA pair by at least 6 fold (Fig. 5.10A). This is because Hfq still acts on the sRNA, and this is sufficient to facilitate annealing with the nascent *rpoS* transcript before its inhibitory secondary structure is formed. Moreover, the action of Hfq variants on *rpoSΔ2* annealing to DsrA and RprA followed different patterns that were consistent with wt *rpoS* annealing to the corresponding sRNAs (Fig. 5.10 A and B compared to Fig. 5.9A and D).

Together, the cotranscriptional annealing assay proposed three different profiles that are signatures of the sRNA-mRNA annealing regions. The annealing activity does not reflect Hfq binding affinity to RNAs; rather it depends on how Hfq restructures the sRNA for base pairing to a specific mRNA. Therefore, annealing profile is a collective result of local secondary structure, complementarity, and the continuity of the complementary regions.

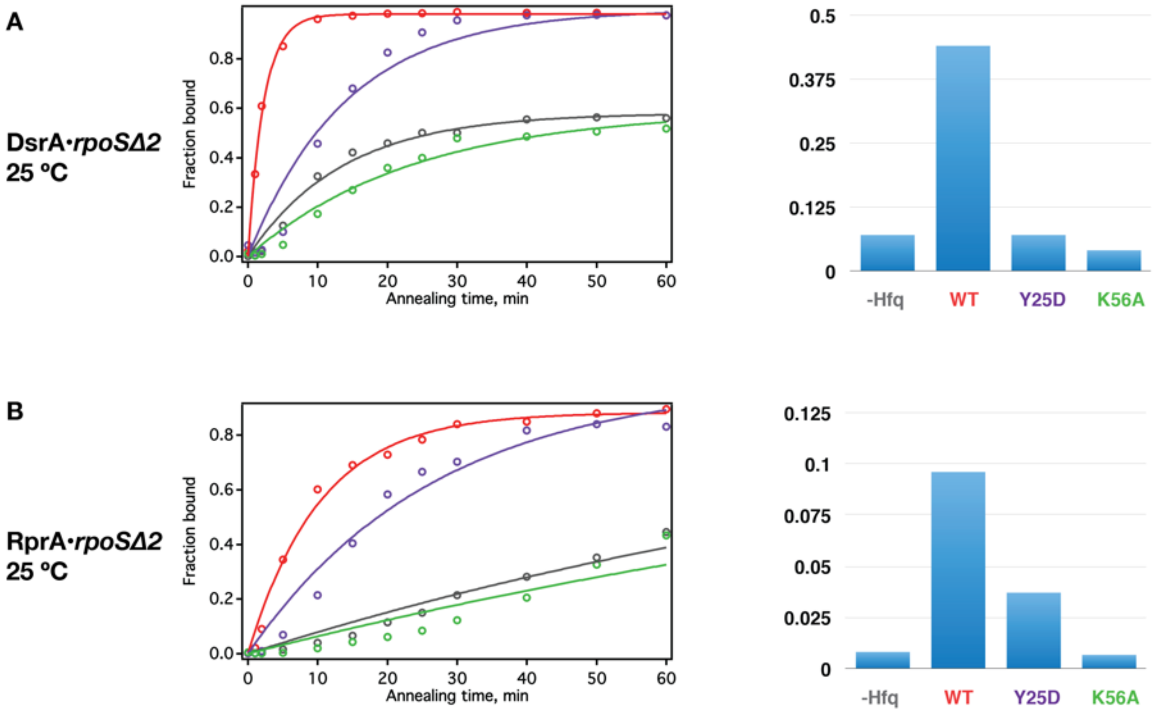


Figure 5.10. Cotranscriptional annealing rates of *rpoSΔ2* and (*A*) DsrA or (*B*) RprA. Left panels, annealing kinetics followed monophasic reaction; right panel, comparison of annealing rates in presence of Hfq variants. Annealing rates were summarized in Table 5.2.

Discussion

I carried out *in vitro* assays to characterize Hfq binding and annealing activity for various sRNAs that contain either A-rich or U-rich motifs. Three different binding modes were observed: the Hfq distal face binds to an internal A-rich motif (ChiX), the Hfq proximal surface binds to the U-tail and the lateral rim binds to the internal U-rich motif (DsrA and McaS), and Hfq-induced structural change in the the sRNA body (RprA). Those various binding modes lead to structural change of sRNAs at various regions. Hfq distal face mutation Y25D showed surprisingly tight binding affinity and promoted higher-order complex formation for all sRNAs; it also induced novel structural change in some cases. One likely explanation is that Y25D changed the Hfq distal face electrostatic properties favoring nonspecific binding so that it adopted a different binding conformation for the sRNAs. This unusual binding behavior may not reflect the typical distal face binding characteristics and additional distal face mutants need to be used for further validation of the preliminary conclusions drawn here.

I used cotranscriptional annealing assays to characterize Hfq's annealing activity for various sRNA-mRNA pairs. The apparent annealing rate may be affected by multiple factors, including transcription rate of the mRNAs, Hfq binding to sRNAs and mRNAs and so on. To reduce the variation due to transcription rate, all the reactions for one sRNA-mRNA pair were prepared and measured simultaneously. Moreover, the transcription rate was unlikely causing a slow annealing rate. In fact, mRNA transcription happened rapidly upon addition of T7-RNAP so that nascent transcripts were able to saturate sRNA in the first 2 min for fast annealing reaction (red and orange curves, Fig. 5.9A and B). As a comparison for the slow annealing reactions, complex formation was significantly lagged in the first 5 min and slowly ramped up later (purple, green, and grey curves, Fig. 5.9A and B). One caveat for the co-transcriptional annealing is that the mRNAs are continuously generated in the reactions, which may cycle off Hfq from the sRNAs that do not anneal efficiently, making the annealing reactions appear more slowly. Finally, this cotranscriptional annealing rate unlikely depends on Hfq•mRNA interaction as discussed previously (Fig. 5.10).

The co-transcriptional annealing profile suggests that various sRNAs behave differently and that even for a given sRNA, different mRNAs may show differential sensitivity to Hfq mutations. This is highly consistent with the in vivo observations (Zhang et al., 2013), although the effects of Hfq mutations on certain RNA pairs were not the same for in vivo and in vitro assays. This is probably because in vivo functional studies are prone to a wide range of factors such as RNA expression and turnover, which was not captured by the in vitro assays. Nevertheless, the preliminary data presented here provided invaluable insight for further characterizing Hfq binding modes and annealing mechanisms for variety of sRNA and mRNA targets.

Materials and methods

Gel mobility shift assays

Preparation of ^{32}P-labeled RNA and gel mobility shift assays were done as previously described (Primers for generating the constructs were listed in Appendix II) (Peng et al., 2014; Soper and Woodson, 2008). For Hfq binding assays, ~70 nM labeled sRNA was combined with 2 µL 5X TNK buffer (50 mM Tris-HCl pH 8.0, 250 mM NaCl, 250 nM KCl), 2 µL 5X Hfq buffer (50 mM Tris–HCl pH 7.5, 1 mM EDTA, 250 mM NH_4Cl) or 2 µL Hfq at various concentrations (0-10 µM monomers) and 2 µL TE in 10 µL total volume and incubated for 10 min at 25˚C. 2 µL aliquots were loaded on a native 6% polyacrylamide gel in 1X TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA). The fractions of bound RNA were quantified as previously described (Lease and Woodson, 2004) and were fit with a partition function assuming one or two binding sites (Chapter 2 eq. 2.1-2.2) (IGOR Pro, WaveMetrics) (Peng et al., 2014).

For co-transcriptional kinetics experiments were done as previously described (Chapter 3 Materials and Methods), 40 µl transcription reaction contained 0.5 mM NTPs, 12.5 ng/µl DNA template, ~50 nM ^{32}P-labeled sRNA with or without 0.6 µM Hfq or 0.25 µM CsrA monomers in 1X T7 RNAP buffer. The mixture was incubated at 37 ºC for 10 min before adding 1 µl T7 RNA polymerase to start transcription. 2 µl reaction was loaded in the native polyacrylamide gel after different times. The fractions bound were quantified and fit with a monophasic annealing equation as previously described (Chapter 2 eq. 2.3) (Peng et al., 2014). Some reactions were adjusted by a lag time term (eq. 5.1) or an amplitude term (eq. 5.2).

$$f_{RD+RDH} = f(1 - \exp(-k_{obs1}(t - t_0))) \quad (5.1)$$

$$f_{RD+RDH} = f_0 + f(1 - \exp(-k_{obs1}t)) \quad (5.2)$$

in which f is the fraction of the reaction that followed the corresponding annealing rate, t is the annealing time, k_{obs1} is the apparent annealing rate for the reaction, t_0 is lag time, f_0 is the amplitude term for adjustion.

RNase footprinting

5’-^{32}P-labeled RNA was prepared as previously described (Lease and Woodson, 2004). 10 µl reaction was assembled as described above containing 1 µl of 0.5 µM labeled sRNA, 1 µl of 5 µM of cold sRNA and 2 µl of 5X Hfq buffer or 15 µM Hfq monomers in 5X Hfq buffer. Reactions were incubated at 37 ºC for 10 min before adding 1 µl of RNase (1 U/ µl RNase If, 1 U/ µl RNase T1, or 0.01 ng/ µl RNase A, concentration determined from dose response as previously described (Peng et al., 2012)) and incubated for another 1 min followed by phenol-chloroform purification. Precipitated sRNAs were resuspended in TE buffer and loading dye for sequencing polyacrylamide gel electrophoresis. Sequences were determined by G-nucleotide from RNase T1 digestion and OH-ladder.

Chapter 6 Discussion

Multi-surface binding mode of Hfq and RNAs

While Hfq has been known to bind A-rich and U-rich sequences on its distal and proximal surfaces since the crystal structures were solved (Link et al., 2009; Schumacher et al., 2002), multi-surface binding mode of Hfq was proposed recently for sRNAs (Sauer, 2013). This thesis research provides the first direct evidence that Hfq lateral surface binds to a U_5 motif in the *rpoS* mRNA, suggesting a new model where Hfq simultaneously contacts the upstream and downstream domains in the long mRNA leader.

I used a variety of *in vivo* and *in vitro* assays to show that the multi-surface binding mode is important for Hfq's function. Disrupting Hfq binding at the U_5 motif reduced Hfq regulation for sRNA-mediated stranslation initiation *in vivo* and sRNA annealing *in vitro*. Moreover, the distance and orientation of the upstream $(AAN)_4$ and downstream U_5 motif in the *rpoS* leader is optimal for Hfq binding and sRNA annealing. Relocating A-rich motif or swapping the $(AAN)_4$ and U_5 motif impaired Hfq function to various degrees. All these evidence suggest that Hfq multiple surfaces simultaneously interact RNA A- and U-rich motifs in a precise manner to facilitate its chaperoning function.

Despite the well-studied molecular rules for the *rpoS* leader, multi-surface binding mode has yet been extensively tested in other mRNA leaders. For most mRNA leaders that are negatively regulated by sRNAs are short (usually <200 nt) and contain an A-rich motif that binds Hfq distal face (Vogel and Luisi, 2011). Only a handful of mRNA leaders form an extensive and complex secondary structure with multiple A- and U-rich patches, including *rpoS*, *flhA*, *cirA*, and *flhD* mRNAs. The *flhA* mRNA leader was proposed to contact both the distal and proximal surfaces of Hfq (Salim and Feig, 2010) whereas *cirA* mRNA directly binds to Hfq at an upstream (AAN) motif and two downstream U-rich patches (Salvail et al., 2013).

By contrast, I found that *flhD* leader only contains one Hfq binding site, the (AAN) motif. The lack of a second Hfq binding site has profound implications for Hfq regulation. Hfq does not effectively alter *flhD* secondary structure and only enhances sRNA annealing moderately. Hfq brings ArcZ sRNA into proximity to the *flhD* leader and restructures McaS sRNA for base pairing.

Clearly, multi-surface binding mode may be widely applied to many Hfq-binding RNAs that contain A-rich and U-rich sequences. Simultaneously contacting multiple surfaces of Hfq not only facilitates RNA restructuring, but also provides a platform for concurrent RNAs to anneal.

Hfq folds RNA tertiary conformation

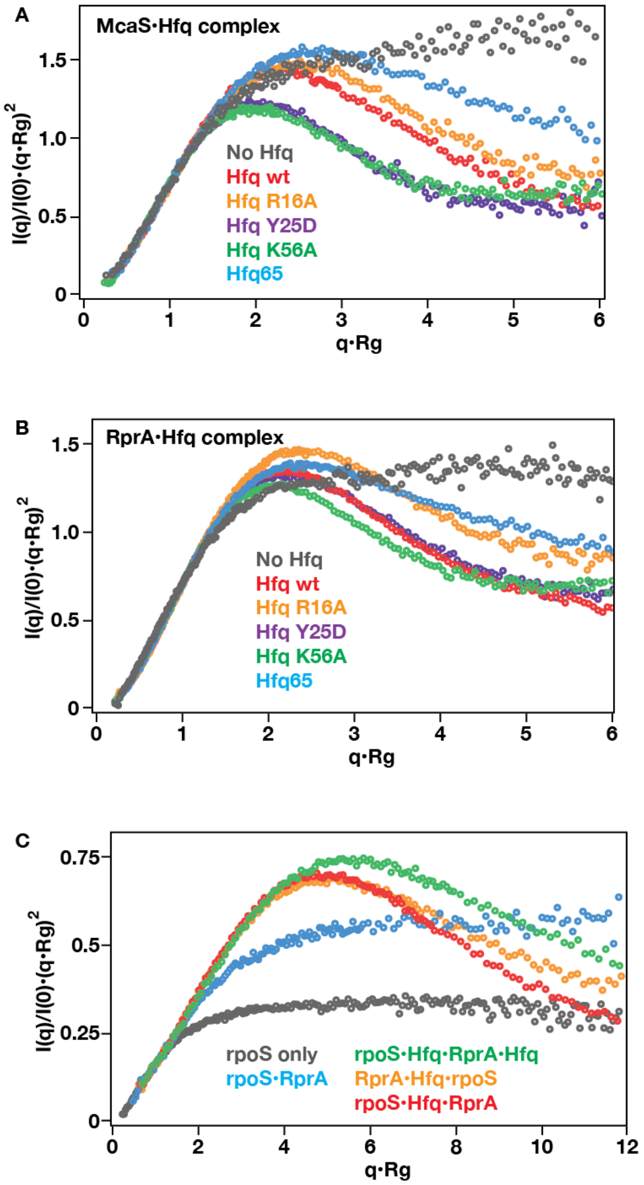
Much effort has been focused on Hfq's chaperoning activities that melt RNA secondary structure for effect interaction with other RNAs or protein factors (Vogel and Luisi, 2011). This thesis research, however, for the first time explored the tertiary conformational change of a long mRNA leader upon Hfq binding in great details. Small angle X-ray scattering data clearly showed that Hfq folds *rpoS* leader into a compact conformation, which was partially disturbed by disrupting Hfq lateral surface binding to the U_5 motif, suggesting that multi-surface binding contributes to this folding effect.

All-atom structural models presented here uncovered invaluable insights to explain how Hfq accomplishes chaperoning activity by folding the *rpoS* mRNA. First, Hfq positions the inhibitory stem over the proximal surface that is known to bind the sRNA. Second, Hfq partially unwinds the inhibitory stem for efficient base pairing and ribosome entry. Third, the tertiary conformation is rather dynamic that releases the downstream domain of *rpoS* RNA for translation initiation.

This folding mechanism also applies to some other Hfq-binding RNAs. For instance, low-resolution *ab inito* models showed that RprA and OxyS sRNAs were wrapped around Hfq (Henderson et al., 2013), in agreement with a distorted tertiary conformation proposed here. Moreover, I obtained SAXS data for McaS and RprA in complex with WT Hfq and Hfq mutants, which showed a more compact conformation in the dimensionless Kratky plot as observed for *rpoS* mRNA (Fig. 6.1A and B). Intriguingly, various Hfq mutations altered the shape of the curves differently, inferring that Hfq may engage multiple surfaces to restructure RNA tertiary conformation. This fully agrees with my model that multi-surface binding is prerequisite for Hfq's folding ability.

I also discovered that Hfq not only folds free RNAs but also RNA duplex (Fig. 6.1C). In the dimensionless Kratky plog, free rpoS and rpoS•RprA displayed a plateaued curve corresponding to an unfolded or flexible conformation (Fig. 6.1C grey and blue curves). By contrast, Hfq folds the RNA duplex into a more compact conformation. Moreover, preforming complex of Hfq with either *rpoS* or RprA RNA generates effectively identical ternary conformation (Fig. 6.1C red and orange curves). Quaternary complex containing both RNAs and two Hfq hexamers also adopted a compact conformation, although differed from the ternary complex as expected (Fig. 6.1C green curve). This compact conformation of higher order complexes was also observed for *rpoS* mRNA binding to Hfq at molar ratio > 2:1. Further study is necessary to investigate the molecular details of those complexes.

Figure 6.1. Hfq folds RNA tertiary conformation in solution. Dimensionless Kratky plot (Durand et al., 2010; Receveur-Brechot and Durand, 2012) of Hfq in complex with (A) McaS sRNA, (B) RprA sRNA, and (C) *rpoS*•RprA duplex (Chapter 4 Materials and Methods). (A,B) sRNA alone is flexible in solution (grey); WT Hfq and Hfq mutants all folds the RNA into compact conformation with slightly different overall shape (red, orange, purple, green, and blue colors). (C) freel *rpoS* mRNA (grey) or *rpoS*•RprA duplex are unfolded (grey and blue); preforming Hfq complex with either RprA (orange) or *rpoS* (red) generated almost identical compact conformation upon ternary complex formation; complex (green) containing the *rpoS* leader, RprA RNA, and 2 Hfq hexamers also formed a compact tertiary structure.


A
McaS·Hfq complex
I(q)/I(0)·(q·Rg)^2
q·Rg
No Hfq
Hfq wt
Hfq R16A
Hfq Y25D
Hfq K56A
Hfq65
0
0.5
1.0
1.5
0
1
2
3
4
5
6
B
RprA·Hfq complex
I(q)/I(0)·(q·Rg)^2
q·Rg
No Hfq
Hfq wt
Hfq R16A
Hfq Y25D
Hfq K56A
Hfq65
0
0.5
1.0
1.5
0
1
2
3
4
5
6
C
I(q)/I(0)·(q·Rg)^2
q·Rg
rpoS only
rpoS·RprA
rpoS·Hfq·RprA·Hfq
RprA·Hfq·rpoS
rpoS·Hfq·RprA
0
0.25
0.5
0.75
0
2
4
6
8
10
12

Variation of Hfq's chaperoning action

Hfq not only displays distinct binding modes for RNA targets, it also employs various mechanisms to facilitate RNA annealing, including "bridging", positioning, restructuring, and stabilizing the RNA duplex (Vogel and Luisi, 2011). By studying a diversity of Hfq-binding mRNAs and sRNAs, I observed great variations of Hfq binding modes and chaperoning activities.

First, Hfq recognizes A- and U-rich sequences in both mRNAs and sRNAs. The classic Hfq-RNA interaction model postulated that Hfq distal face binds A-rich sequence in the mRNA while the proximal face binds U-rich sequence in the sRNA(Link et al., 2009; Schumacher et al., 2002; Wang et al., 2013). I found, however, U-rich motifs in *rpoS* but not *flhD* mRNA body directly contact the arginine patch on the lateral rim. I also discovered that an A-rich sequence in ChiX sRNA directly contacts Hfq distal face, leaving the 3' U-tail dispensable for tight Hfq binding. Some sRNAs may adopt more complicated Hfq binding mode. For instance, RprA sRNA, previously shown to bind both proximal and distal surfaces (Updegrove and Wartell, 2011), does not require 3' U-tail for tight Hfq binding but lacks a profound A-rich motif for distal face interaction.

Another binding variation stems from Hfq C-terminus, which was previously shown to be required for long mRNAs binding (Vecerek et al., 2008). Surprisingly, a C-terminal truncation Hfq65 binds to sRNAs of similar length very differently (Fig. 6.2). Hfq65 does not form a stable complex with McaS sRNA while binding with DsrA sRNA nicely during native polyacrylamide gel electrophoresis, even though the two sRNAs share similar secondary structure complexity and both require 3' U-tail for tight binding to Hfq proximal surface. Moreover, Hfq65 does not complex with ChiX sRNA well in the gel, despite that ChiX sRNA binds to all other Hfq mutants tightly. Although I cannot fully explain these observations yet, one possibility is that Hfq C-terminus prevents nonspecific RNA binding by covering Hfq distal and lateral surfaces. This is consistent with my structural model of the full-length Hfq showing frequent occupancy of C-terminus over those surfaces.

Lastly, Hfq employs distinct annealing mechanisms for the same RNA with different base pairing partners. For instance, Hfq mainly “bridges” ArcZ56 sRNA while actively restructures McaS sRNA for *flhD* mRNA annealing. In the same line, Hfq engages different surfaces to facilitate co-transcription annealing of *flhD* and *csgD* mRNAs to McaS sRNA. I propose that a co-transcriptional annealing profile can be used to distinguish requirement of various Hfq surfaces, which correlates with the complementarity of mRNA-sRNA pairs and structural complexity of the annealing RNAs.

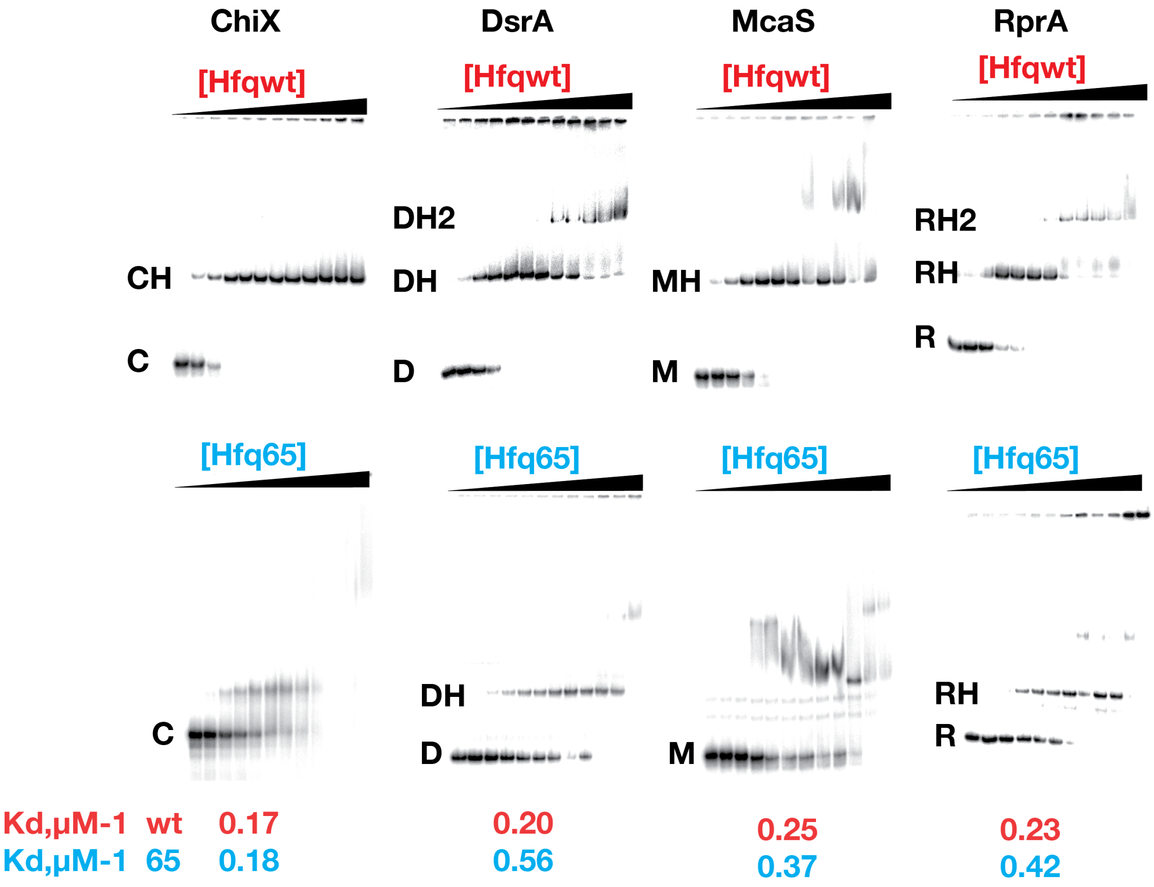


Figure 6.2. Comparing WT Hfq and Hfq65 binding with sRNAs. WT Hfq and Hfq65 titrations (concentration ranges from 0-2 µM monomers, Chapter 5 Materials and Methods) of uniformly labeled ChiX, DsrA, McaS, and RprA sRNAs at 25 ºC. Free sRNA (lower band) binds one or two Hfq multimers (first and second band shifts). Fraction bound was fit with a partition equation assuming one specific binding site (Chapter 2 Materials and Methods eq. 2.1). Binding constants were labeled at the bottom of the gels.

Conclusions

At the completion of this thesis research, we have gained a much more comprehensive view of Hfq's chaperoning mechanism as exemplified by *rpoS*•DsrA annealing (Fig. 6.3). Hfq initially binds the A-rich and U-rich motifs in the mRNA by its distal face and the lateral surface, forming a closed complex that positioned the sRNA annealing site over the proximal surface and partially unwound the inhibitory stem. The incoming sRNA 3'-U tail anchors at the proximal pocket while the sRNA body binds the lateral surface. This close proximity between the seeding region in the sRNA and the opened annealing site in the mRNA leads to effective duplex nucleation, which is a thermodynamic process with no known assistance from Hfq. Finally, the mRNA•sRNA duplex zips off the proximal and the lateral surfaces while remaining at the distal face via the A-rich motif in the mRNA. The resulting open complex is ready to cycle off the RNA duplex at the presence of other competing RNAs.

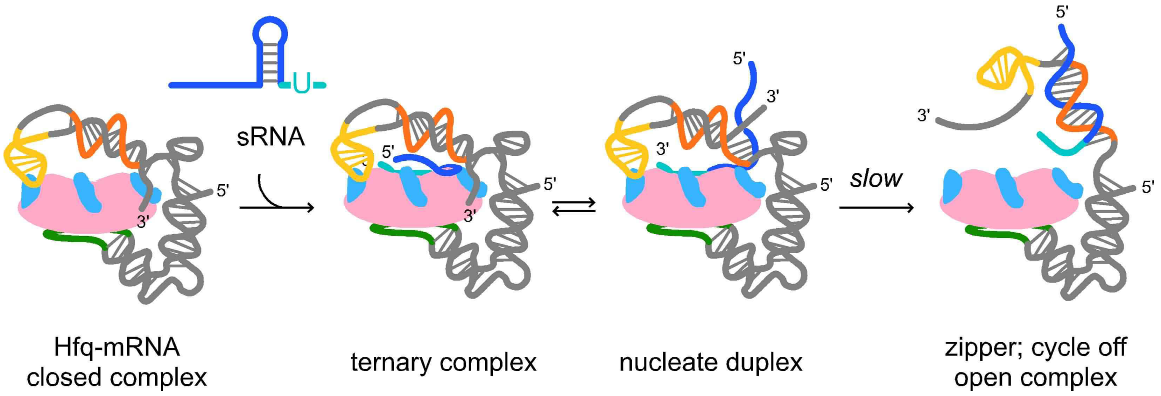


Figure 6.3. Model of Hfq chaperoning activity (courtesy of Dr. Sarah A. Woodson). Side view of Hfq (pink) shows the lateral binding site (light blue) and the proximal surface facing up. The mRNA leader (grey) has complex secondary structure with an A-rich motif (dark green), a U-rich motif (gold), and the sRNA annealing site (orange). The sRNA (dark blue) has a simple stem-loop structure with internal U-rich patch and a U-rich tail (light green).

Future perspectives

Hfq is a fascinating global regulator that controls a vast variety of RNA-related activities in bacterial cells. In the past few years, tremendous progress has been made in the Hfq community, which greatly enriches our knowledge of Hfq protein and its constellation of RNA targets.

One major direction is to further understand structural features and functional implications of Hfq proteins. Horizontal comparison of Hfq homologs among bacterial species and that of Hfq to archaeal/eukaryotic Sm/Lsm proteins revealed conservation of critical structural elements as well as variation of newly adapted features. For instance, conserved acidic and basic residues on Hfq surfaces may establish electrostatic landscape for RNA interactions. Different C-terminus extensions may interact with a subset of RNAs that are present in specific bacterial species.

Another exciting field is to characterize and classify Hfq-interacting RNAs. For Hfq binding mode, three lines of directions include: how RNAs present their A- and U-rich motifs for Hfq recognition in the context of a secondary structure; how RNAs coordinate multiple Hfq binding surfaces to accommodate secondary and tertiary conformations; how various RNAs engage Hfq C-terminus for specific interaction. For Hfq activity, much more research is needed to explore detailed mechanisms for canonical and non-canonical functions.

Appendix I

Solutions preparation

Acrylamide/Bis-acrylamide, 40% solution

1 day before: bake 2L beaker with a mixer bar, 1L grad cylinder, 2L side-arm flask, funnel, spatulas and 1L dark storage bottle. Wear face mask, gloves, goggles and lab coat. Work with caution and post a warning note.

Acrylamide	386.6 g
Bisacrylamide	13.34 g

Carefully measure the powder and decant in the beaker. Add ~600ml RNase-free H_2O and mix the solution well on a magnetic stage. Wrap the beaker with aluminum foil to block the light.

RNase-free H_2O	Bring to 1L

Filter the solution. Keep at 4 ºC for up to 6 months.

APS, 10% solution

Ammonium persulfate (MW 228.2)	10g
RNase-free H_2O	Bring up to 100ml

Keep the solution at 4ºC for up to 6 months.

Ampicillin stock solution

100 mg/ml in H_2O. Dilute as 1:1000 into LB when use.

Chloramphenicol stock solution

25mg/ml in 100% EtOH. Dilute as 1:1000 into LB when use.

DTT, 1 M solution

DTT (MW 154.25)	0.771g
RNase-free H_2O	Bring up to 5ml

Filter the solution with nylon syringe filter (0.22 µm). Make 1ml aliquots and store at -20 ºC for up to a year. Do not freeze-thaw the aliquots for more than 3 times.

EDTA, 0.5 M solution

EDTA dihydrate, (MW 372.2)	18.61g

Add H_2O to about 80ml. Slowly add 6M NaOH by increment of 0.5ml until EDTA is dissolved. Measure pH with pH test strip and adjust to pH=8.0. Add additional H_2O to bring volume to 100ml.

Autoclave after preparation. Keep at room temperature for up to 1 year.

IPTG, 1 M solution

IPTG (MW 238.3)	2.38g
RNase-free H_2O	Bring up to 10ml

Filter the solution with nylon syringe filter (0.22 µm). Make 1ml aliquots and store at -20 ºC for up to a year. Do not freeze-thaw the aliquots for more than 3 times.

KCl, 3 M solution

KCl (MW 74.55)	22.35g
RNase-free H_2O	Bring up to 100ml

Autoclave after preparation. Keep at room temperature for up to 3 year.

$MgCl_2$, 1 M solution

$MgCl_2$•$6H_2O$ (MW 203.3, hygroscopic and kept in desiccator)	20.33g
RNase-free H_2O	Bring up to 100ml

Autoclave after preparation. Keep at room temperature for up to 3 year.

$MnCl_2$, 1 M solution

$MnCl_2$•$4H_2O$ (MW 197.9)	19.79g
RNase-free H_2O	Bring up to 100ml

Autoclave after preparation. Keep at room temperature for up to 3 year.

NaCl, 4 M solution

NaCl (MW 58.44)	23.4g
RNase-free H_2O	Bring up to 100ml

Autoclave after preparation. Keep at room temperature for up to 3 year.

NaOAc, 3 M pH 5.2 solution

NaOAc (MW 82.03)	24.6g

Dissolve the solutes in 80ml H_2O. Adjust pH to 5.2 with acetic acid.

RNase-free H_2O	Bring up to 100ml

Autoclave after preparation. Keep at room temperature for up to 3 year.

NaOH, 6 M solution

NaOH pellet (MW 40)	24g
RNase-free H_2O	Bring up to 100ml

Autoclave after preparation. Keep at room temperature for up to 3 year.

NH_4Cl, 1 M solution

NH_4Cl (MW 53.5)	5.35g
RNase-free H_2O	Bring up to 100ml

Autoclave after preparation. Keep at room temperature for up to 3 year.

Tris-HCl, 1 M pH 8.0 solution

THAM (MW 121.1)	12.11g

Dissolve in 80ml H_2O. Adjust pH to 8.0 with HCl.

RNase-free H_2O	Bring up to 100ml

Autoclave after preparation. Keep at room temperature for up to 1 year.

SDS, 10% solution

SDS (MW 288.38)	10g

Add RNase-free H_2O to ~90ml. Heat the solution at 65ºC while shaking occasionally until SDS is completely dissolved. Cool down the solution to room temperature.

RNase-free H_2O	Bring up to 100ml

No need to sterilize. Keep at room temperature for up to 3 year.

Spermidine, 1M solution

Spermidine (MW 145.25)	0.726g
RNase-free H_2O	Bring up to 5ml

Filter the solution with nylon syringe filter (0.22 µm). Make 1ml aliquots and store at -20 ºC for up to a year. Do not freeze-thaw the aliquots for more than 3 times.

Buffer preparations

5X Hfq storage buffer

Ingredients	Amount	Final concentration
Tris pH 7.5 (1M)	500µl	50mM
EDTA (0.5M)	20µl	1mM
NH_4Cl (1M)	2.5ml	250mM
100% glycerol	1ml	10% v/v
RNase free H_2O	Bring to 10ml	

10X PNK buffer

Ingredients	Amount	Final concentration
Tris pH 7.5 (1M)	7ml	700mM
$MgCl_2$ (1M)	1ml	100mM
DTT (1M, less than 6 months old)	0.5ml	50mM
RNase free H_2O	Bring to 10ml	

RNA elution buffer

Ingredients	Amount	Final concentration
SDS (10%)	500µl	0.1%
EDTA (0.5M)	100µl	1mM
NaOAc (3M)	8.3ml	0.5M
RNase free H_2O	Bring to 50ml	

SAXS sample storage buffer

Ingredients	Amount	Final concentration
Tris (121.14)	6.057g	50mM
NaCl (58.44)	2.922g	50mM
KCl (74.55)	3.728g	50mM
NH4Cl (53.49)	2.674g	50mM
MgCl2 (1M)	2ml	2mM
glycerol	20ml	2%

Adjust pH to 7.5

RNase free H_2O	Bring to 1L

10X T7 RNAP buffer

Ingredients	Amount	Final concentration
Tris pH 8.0 (1M)	4ml	400mM
$MgCl_2$ (1M)	1.5ml	150mM
DTT (1M, less than 6 months old)	500ul	50mM
spermidine (1M)	200ul	20mM
RNase free H_2O	Bring to 10ml	

10X TBE buffer

Ingredients	Amount	Final concentration
THAM (MW 121.1)	108g	890mM
EDTA (0.5M)	40ml	20mM
Boric acid (MW 61.83)	55g	890mM
RNase free H_2O	Bring to 1L	

TE buffer

Ingredients	Amount	Final concentration
Tris pH 7.5 (1M)	100µl	10mM
EDTA (0.5M)	20µl	1mM
RNase free H_2O	Bring to 10ml	

TEN buffer

Ingredients	Amount	Final concentration
Tris pH 7.5 (1M)	500µl	10mM
EDTA (0.5M)	100µl	1mM
NaCl (4M)	3.125ml	250mM
RNase free H_2O	Bring to 50ml	

10X THEM2 buffer

Ingredients	Amount	Final concentration
HEPES (MW 238.3)	157.07g	660mM
Tris aminomethane (MW 121.1)	41.14g	340mM
EDTA (MW 292.2)	0.29g	1mM
$MgCl_2 \cdot 6H_2O$ (MW 203.3)	4.06g	20mM
RNase free H_2O	Bring to 1L	

5X TNK buffer

Ingredients	Amount	Final concentration
Tris pH 7.5 (1M)	0.5ml	50mM
NaCl (4M)	0.625ml	250mM
KCl (3M)	0.833ml	250mM
RNase free H_2O	Bring to 10ml	

Electrophoresis loading dye preparations

Glycerol dye (for native polyacrylamide gel)

Ingredients	Amoung
100% glycerol	500ul
BP dye (2%)	50µl
XC dye (2%)	20µl
RNase-free H_2O	430ul

For native polyacrylamide gel electrophoresis. Use as 10X loading dye for samples (up to 40X for co-transcriptional annealing assays). Store at 4 ºC for up 6 months.

Formamide dye (for sequencing polyacrylamide gel)

Ingredients	Amoung
De-ionized formamide (0.5g AG501-X8 resin beads equilibrated in 10ml formamide for >30min)	9.5ml
10X TBE	400µl
XC dye (2%)	50µl
BP dye (2%)	50µl
Make 1ml aliquots.	

For sequencing polyacrylamide gel electrophoresis. Use as resuspending buffer for desiccated samples (or up to 3X loading dye for solution samples). Make 1 ml aliquots and store at -20 ºC for up 1 year.

5X SDS-PAGE loading dye

Ingredients	Amoung
0.5M Tris pH6.8	1.25ml
Glycerol	2ml
10% SDS	2ml
β-mercaptoethanol	0.5ml
BP dye (2%)	0.5ml
RNase-free H_2O	bring to 10ml

For SDS-PAGE electrophoresis. Use as 5X loading dye for protein solution samples. For denaturing gel electrophoresis, boil the samples in the loading buffer at 95ºC for 5min. For semi-native gel electrophoresis, load the sample in buffer without heat treatment. Store at room temperature for up to 1 year.

10X TAE dye

Ingredients	Amoung
TAE (1X)	50μl
Sucrose	1g
BP dye (2%)	250μl
XC dye (2%)	62.5μl
RNase-free H_2O	637.5μl

For agarose gel electrophoresis. Use as 10X loading dye. Store at room temperature for up to 3 year.

2X TBE dye

Ingredients	Amoung
Urea	6.006g
10X TBE	1ml
Keep at 65ºC water bath to dissolve urea	
BP dye (2%)	200µl
XC dye (2%)	200µl
RNase-free H_2O	bring to 10ml

For denaturing polyacrylamide gel electrophoresis. Use as 2X loading dye for solution samples (or up to 5X). Store at room temperature for up to 1 year.

Appendix II

Primers for generating rpoS-A18 insertions

53A18F	AAAAAAAAAAAAAAAAAACGGTCACAGCGCCTGTAACGG
53A18R	GGTCGCAGTTGGCTTGTTGTTCGGC
250A18F	AAAAAAAAAAAAAAAAAATGGCCGCGTTGTTTATGCTGGTAACG
250A18R	GCGGTCGCGATAATTGCCTGTCCTTTGC
366A18F	AAAAAAAAAAAAAAAAAAGTTAAGGCGGGGCAAAAAATAGCG
366A18R	GGACCAGCATTGTGTCGTTATGG
441A18F	AAAAAAAAAAAAAAAAAATTTGAAATTCGTTACAAGGGG
441A18R	GCAAGCGTGTTGAACTGGTTCC
484A18F	AAAAAAAAAAAAAAAAAAGCCGCAGCGATAAATCGGCGG
484A18R	ACGCAGCGGGTTTACGGATTTCCCC
499A18F	AAAAAAAAAAAAAAAAAATCGGCGGAACCAGGCTTTTGC
499A18R	CGCTGCGGCAAATAACGCAGCGGG
519A18F	AAAAAAAAAAAAAAAAAAGCTTGAATGTTCCGTCAAGGG
519A18R	GCCTGGTTCCGCCGATTTATCGC

Primers for generating Pbad-rpoS-lacZ PCR fragments

Pbad-rpoS254F	ACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATGGCCGCGTTGTTTATGCTGG
Pbad-rpoS401F	ACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATGCGACCATGGGTAGCACCGG
Pbad-rpoS439F	ACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATGCATTTTGAAATTCGTTACA
Pbad-rpoS439A18F	ACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATAAAAAAAAAAAAAAAAAAGC
Pbad-rpoS420A18F	ACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATAAAAAAAAAAAAAAAAAAGAACCAGTTCAACACGC
lacZ-rpoS10aaR	TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACATCATGAACTTTCAGCGTATTCTGACTCATAAGGTGGCTCC

Primers for generating rpoS mutations

rpoS-U5mut_F	CCCCCGCCGCAGCGATAAATCGGCGG
484A18R	ACGCAGCGGGTTTACGGATTTCCCC
rpoS-U4_F	CCCCGCTTGAATGTTCCGTCAAGGG
rpoS-U4_R	GCCTGGTTCCGCCGATTTATCGCTGCG
rpoS-UA3_F	ATGCGTCGGCGGAACCAGGCTTTTGC
rpoS-UA3_R	CGCTGCGGCAAATAACGCAGCGGG
rpoS-AU4_F	GCATGCTGAAATTCGTTACAAGGGGAAATCCGTAAACCC
rpoS-AU4_R	GCGTGTTGAACTGGTTCCGGTGC
rpoS-U5LS_F	GCAGGCTTATTTGCGCCTGCCGCAGCGATAAATCGGCGGAACC
rpoS-U5UL_F	GCCTCGAGCTTATTTCGCAGCGATAAATCGGCGGAACCAGGC
rpoS-U5DL_F	TTATTTCCGAGTCGCCGCAGCGATAAATCGGCGGAACCAGGC
rpoS-U5LS_R	CGCAGCGGGTTTACGGATTTCCCC
rpoS-U5SS_F	GCGTTATTTCGCGATAAATCGGCGGAACCAGGCTTTTGC
rpoS-U5SS_R	GGGTTTACGGATTTCCCCTTGTAACG
rpoS-369U5_F	CCGCCTTATTTGCGGCGGCGTTAAGGCGGGGCAAAAAATAGCG
rpoS-369U5_R	CCCGGACCAGCATTGTGTCGTTATGG

Primers for generating rpoS templates for in vitro transcription

T7-rpoS323F	GTG**TAATACGACTCACTATA**GGCCGCGTTGTTTATGCTGG
T7-rpoS301F	GTG**TAATACGACTCACTATA**GGGTGAACGATTATCATCAAACATAATG
T7-rpoS295F	GTG**TAATACGACTCACTATA**GGGTGAACAGAGCAATAATCTGATTATCATCAAAC
T7-rpoS419GF	GTG**TAATACGACTCACTATAG**GGAACCAGTTCAACACGCTTGC
T7-rpoS439G2F	GTG**TAATACGACTCACTATAGG**GCATTTTGAAATTCGTTACAAGG
T7-rpoS323R	AATTCTGACTCATAAGGTGGCTCCTACC
T7-rpoS596R	AAATCATGAACTTTCAGCGTATTCTGACTCATAAGG
T7-rpoSupF	GTG**TAATACGACTCACTATA**GGGTGCTACCTGAGTGCCTACGCCCATAACGACAC
T7-rpoSupR	GGGTGCTACCCATGGTCGCTATTTTTGCCCC

Primers for generating rpoS cDNA fragments by reverse transcription

PE358	ACTTCTTGTTGTTCCCGGACCAGC
PE467	AAATAACGCAGCGGGTTTACGG
PE418	GCAAGCGTGTTGAACTGGTTCCG
PE477	TATCGCTGCGGCAAATAACGCAGC
PE508	CATTCAAGCAAAAGCCTGGTTCCG
PE554	AATTCTGACTCATAAGGTGGCTCC

Primers for generating flhD templates for in vitro transcription

T7-flhD_F	GTG**TAATACGACTCACTATA**GATTTAGGAAAAATCTTAGATAAGTGTAAAGACCCATTTCTATTTGTAAGG
T7-flhD3G_F	GTG**TAATACGACTCACTATAGG**GATTTAGGAAAAATCTTAGATAAGTGTAAAGACCCATTTCTATTTGTAAGG
T7-flhD6_R	ATGCATTATTCCCACCCAGAATAACC
T7-flhD9_R	GCAACTCGGAGGTATGCATTATTCCC
T7-flhD14_R	TCGGAGGTATGCATTATTCCC
T7-flhD27_R	GTGTTTCAGCAACTCGGAGGTATGCATTATTCCCACCCAGAATAACC

Primers for cloning flhD into pUC18-T7 plasmids

HindIII_T7_flhD_F	GC**AAGCTTTAATACGACTC**
EcoRI_flhD_R	GC**GAATTC**GTGTTTCAGCAACTCGG

Primers for generating flhD cDNA fragments by reverse transcription

PE27R	GTGTTTCAGCAACTCGGAGG

Primers for generating sRNAs for in vitro transcription

T7-ChiX_F	GTG**TAATACGACTCACTATAGGG**ACACCGTCGCTTAAAGTGACGGC
ChiX_R	AAAAAAATGGCCAATATCGCTATTGGCCCG
ChiX_delU_R	TGGCCAATATCGCTATTGGCCCG
T7-DsrA_F	GTG**TAATACGACTCACTATAGGG**AACACATCAGATTTCCTGGTGTAACG
DsrA_R	AAAAAAATCCCGACCCTGAGGGGGTCGGG
DsrA_delU_R	TCCCGACCCTGAGGGGGTCGGG
T7-McaS_F	GTG**TAATACGACTCACTATAGGG**ACCGGCGCAGAGGAGACAATGCCGG
McaS_R	AAAAAAATAGAGTCTGTCGACATCCGCC
McaS_delU_R	TAGAGTCTGTCGACATCCGCC
T7-RprA_F	GTG**TAATACGACTCACTATAGGG**ACGGTTATAAATCAACATATTGATTTATAAGC
RprA_R	AAAAAAAGCCCATCGTGGGAGATGGG
RprA_delU_R	GCCCATCGTGGGAGATGGGC
T7-ArcZ56_F	GTG**TAATACGACTCACTATAGGG**TTTCCCTGGTGTTGGCGCAGTATTCG
ArcZ56_R	AAAAAATGACCCCGGCTAGACCG

Primers for generating mRNA leaders for in vitro transcription

T7-ChiP266_F	GTG**TAATACGACTCACTATAGGG**CGCTTTCATCTTGCTGGAATTTCACGCTTTTATTC
T7-ChiP100_F	GTG**TAATACGACTCACTATAGGG**CTGTAGTCAGCGAGACTTTTCTCAACGCTAC
T7-ChiP10aa_R	ACTACGTTTGCCACTAAACGTACGCATGGGTTAATCCTC
T7-ChiP39aa_R	TGAATCGTCGATGAATCCTTCGGCGCGAGCCTCCGGC
csgD_F	TATAGGGCAGATGTAATCCATTAGTTTATATTTTACCC
csgD_30R	ACCATGAATACTATGGACTTCATTAAACATGATG
pgaA_F	TATAGGGAGGCATTGGGATTTATGCCGTATTCCTGAAG
pgaA_30R	CGGGCACCTTTTTCTGCTACTTGAATACATCCTGTATTAC

Bacterial strains and plasmids

Names	Description
PM1205(1)	MG1655 *mal*::*lacI*^{q}, ΔaraBAD, *lacI*'::PBAD-*cat-sacB:lacZ*, miniλtet^R
TSdelAdel6(2)	PM1205 *lacI*'::PBAD-*rpoS*-Δ(AAN)4ΔA6*:lacZ*
TS576(2)	PM1205 *lacI*'::PBAD-*rpoS*-Δ(AAN)4*:lacZ*
TSdel6(2)	PM1205 *lacI*'::PBAD-*rpoS*-ΔA6*:lacZ*
PY1230	PM1205 *lacI*'::PBAD-*rpoS*-Δ(AAN)4ΔA6-53A18*:lacZ*
PY1231	PM1205 *lacI*'::PBAD-*rpoS*-Δ(AAN)4ΔA6-250A18*:lacZ*
PY1203	PM1205 *lacI*'::PBAD-*rpoS*-Δ(AAN)4ΔA6-366A18*:lacZ*
PY1202	PM1205 *lacI*'::PBAD-*rpoS*-Δ(AAN)4ΔA6-441A18*:lacZ*
PY1201	PM1205 *lacI*'::PBAD-*rpoS*-Δ(AAN)4ΔA6-484A18*:lacZ*
PY1232	PM1205 *lacI*'::PBAD-*rpoS*-Δ(AAN)4ΔA6-499A18*:lacZ*
PY1233	PM1205 *lacI*'::PBAD-*rpoS*-Δ(AAN)4ΔA6-519A18*:lacZ*
PM1552(2)	TS576 *hfq::cat*
PM1556(2)	TSdelAdel6 *hfq::cat*
PY1213	PY1203 *hfq::cat*
PY1212	PY1202 *hfq::cat*

PY1211	PY1201 *hfq::cat*
PY1204	PM1205 *lacI*'::PBAD-*rpoS*-Δ484*:lacZ*
PY1237	PM1205 *lacI*'::PBAD-*rpoS*-Δ499*:lacZ*
PY1208	PM1205 *lacI*'::PBAD-*rpoS*-Δ519*:lacZ*
PY1238	PM1205 *lacI*'::PBAD-*rpoS*-499A18*:lacZ*
PY1239	PM1205 *lacI*'::PBAD-*rpoS*-519A18*:lacZ*
pRpoS576(3)	Topo based template containing full-length rpoS
pRpoSDA6(2)	Topo based template containing full-length rpoS with A6 mutation
pRpoSDAAYAA(2)	Topo based template containing full-length rpoS with (AAN)4 mutation
pRpoS-HBM(2)	Topo based template containing full-length rpoS with A6(AAN)4 mutation
pRpoS-Δ2-53A18	Topo based template containing full-length rpoS with A6(AAN)4 mutation and A18 insertion at position 53
pRpoS-Δ2-250A18	Topo based template containing full-length rpoS with A6(AAN)4 mutation and A18 insertion at position 250
pRpoS-Δ2-366A18	Topo based template containing full-length rpoS with A6(AAN)4 mutation and A18 insertion at position 366
pRpoS-Δ2-441A18	Topo based template containing full-length rpoS with A6(AAN)4 mutation and A18 insertion at position 441
pRpoS-Δ2-484A18	Topo based template containing full-length rpoS with A6(AAN)4 mutation and A18 insertion at position 484
pRpoS-Δ2-499A18	Topo based template containing full-length rpoS with A6(AAN)4 mutation and A18 insertion at position 499

pRpoS-Δ2-519A18	Topo based template containing full-length rpoS with A6(AAN)4 mutation and A18 insertion at position 519
pRpoS-Δ484	Topo based template containing full-length rpoS with U-rich motif mutation at position 484
pRpoS-Δ499	Topo based template containing full-length rpoS with A-rich motif mutation at position 499
pRpoS-Δ519	Topo based template containing full-length rpoS with U-rich motif mutation at position 519
pRpoS-wt-499A18	Topo based template containing full-length rpoS with A18 insertion at position 499
pRpoS-wt-519A18	Topo based template containing full-length rpoS with A18 insertion at position 519
pBRplac(4)	Amp^r, Plac promoter based expression vector
pDsrA(2)	AatII-EcoRI DsrA-containing fragment cloned into pBR-plac
pRprA (2)	AatII-EcoRI RprA-containing fragment cloned into pBR-plac
pArcZ (2)	AatII-EcoRI ArcZ-containing fragment cloned into pBR-plac
pUCT7DsrA-U6	pUC18 based DsrA with extended 3’ U6 template with DraI linearization site

---

(1) Mandin, P. & Gottesman, S. 2009, "A genetic approach for finding small RNAs regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-component system", Molecular microbiology, vol. 72, no. 3, pp. 551-565.

(2) Soper, T., Mandin, P., Majdalani, N., Gottesman, S. & Woodson, S.A. 2010, "Positive regulation by small RNAs and the role of Hfq", Proceedings of the National Academy of Sciences of the United States of America, vol. 107, no. 21, pp. 9602-9607.

(3) Soper, T.J. & Woodson, S.A. 2008, "The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA", RNA (New York, N.Y.), vol. 14, no. 9, pp. 1907-1917.

(4) Guillier, M. & Gottesman, S. 2006, "Remodelling of the Escherichia coli outer

membrane by two small regulatory RNAs", Molecular microbiology, vol. 59, no. 1, pp. 231-247.

Reference

Arluison, V., Derreumaux, P., Allemand, F., Folichon, M., Hajnsdorf, E., and Regnier, P. (2002). Structural Modelling of the Sm-like Protein Hfq from Escherichia coli. J. Mol. Biol. *320,* 705-712.

Baba, S., Someya, T., Kawai, G., Nakamura, K., and Kumasaka, T. (2010). Expression, crystallization and preliminary crystallographic analysis of RNA-binding protein Hfq (YmaH) from Bacillus subtilis in complex with an RNA aptamer. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. *66,* 563-566.

Bandyra, K.J., Said, N., Pfeiffer, V., Gorna, M.W., Vogel, J., and Luisi, B.F. (2012). The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. Mol. Cell *47,* 943-953.

Battesti, A., Majdalani, N., and Gottesman, S. (2011). The RpoS-mediated general stress response in Escherichia coli. Annu. Rev. Microbiol. *65,* 189-213.

Behrouzi, R. (2012). The roles of native tertiary interactions in the folding pathway of the Azoarcus ribozyme. ProQuest Dissertations and Theses

Beich-Frandsen, M., Vecerek, B., Konarev, P.V., Sjoblom, B., Kloiber, K., Hammerle, H., Rajkowitsch, L., Miles, A.J., Kontaxis, G., Wallace, B.A.*, et al.* (2011). Structural insights into the dynamics and function of the C-terminus of the E. coli RNA chaperone Hfq. Nucleic Acids Res. *39,* 4900-4915.

Beich-Frandsen, M., Vecerek, B., Sjoblom, B., Blasi, U., and Djinovic-Carugo, K. (2011). Structural analysis of full-length Hfq from Escherichia coli. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. *67,* 536-540.

Beisel, C.L., Updegrove, T.B., Janson, B.J., and Storz, G. (2012). Multiple factors dictate target selection by Hfq-binding small RNAs. EMBO J. *31,* 1961-1974.

Bilusic, I., Popitsch, N., Rescheneder, P., Schroeder, R., and Lybecker, M. (2014). Revisiting the coding potential of the E. coli genome through Hfq co-immunoprecipitation. RNA Biol. *11,* 641-654.

Blanchet, C.E., and Svergun, D.I. (2013). Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution. Annu. Rev. Phys. Chem. *64,* 37.

Boggild, A., Overgaard, M., Valentin-Hansen, P., and Brodersen, D.E. (2009). Cyanobacteria contain a structural homologue of the Hfq protein with altered RNA-binding properties. FEBS J. *276,* 3904-3915.

Brescia, C.C., Mikulecky, P.J., Feig, A.L., and Sledjeski, D.D. (2003). Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. RNA *9,* 33-43.

Brooks, B.R., Brooks, C.L.,3rd, Mackerell, A.D.,Jr, Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S.*, et al.* (2009). CHARMM: the biomolecular simulation program. J. Comput. Chem. *30,* 1545-1614.

Brown, L., and Elliott, T. (1996). Efficient translation of the RpoS sigma factor in Salmonella typhimurium requires host factor I, an RNA-binding protein encoded by the hfq gene. J. Bacteriol. *178,* 3763.

Butland, G., Peregrin-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N.*, et al.* (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature *433,* 531-537.

Chao, Y., and Vogel, J. (2010). The role of Hfq in bacterial pathogens. Curr. Opin. Microbiol. *13,* 24.

Christiansen, J.K., Nielsen, J.S., Ebersbach, T., Valentin-Hansen, P., Sogaard-Andersen, L., and Kallipolitis, B.H. (2006). Identification of small Hfq-binding RNAs in Listeria monocytogenes. RNA *12,* 1383-1396.

Cohen-Or, I., Shenhar, Y., Biran, D., and Ron, E.Z. (2010). CspC regulates rpoS transcript levels and complements hfq deletions. Res. Microbiol. *161,* 694-700.

Collins, B.M., Harrop, S.J., Kornfeld, G.D., Dawes, I.W., Curmi, P.M., and Mabbutt, B.C. (2001). Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs. J. Mol. Biol. *309,* 915.

Curtis, J.E., Raghunandan, S., Nanda, H., and Krueger, S. (2012). **SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints**. Computer Physics Communications *183,* 382-389.

Dambach, M., Irnov, I., and Winkler, W.C. (2013). Association of RNAs with Bacillus subtilis Hfq. PLoS One *8,* e55156.

Das, R., Laederach, A., Pearlman, S.M., Herschlag, D., and Altman, R.B. (2005). SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. RNA *11,* 344-354.

Davis, B.M., and Waldor, M.K. (2007). RNase E-dependent processing stabilizes MicX, a Vibrio cholerae sRNA. Mol. Microbiol. *65,* 373-385.

de Haseth, P.L., and Uhlenbeck, O.C. (1980). Interaction of Escherichia coli host factor protein with oligoriboadenylates. Biochemistry (N. Y.) *19,* 6138.

de Haseth, P.L., and Uhlenbeck, O.C. (1980). Interaction of Escherichia coli host factor protein with Q beta ribonucleic acid. Biochemistry (N. Y.) *19,* 6146.

De Lay, N., and Gottesman, S. (2012). A complex network of small non-coding RNAs regulate motility in Escherichia coli. Mol. Microbiol. *86,* 524-538.

De Lay, N., Schu, D.J., and Gottesman, S. (2013). Bacterial small RNA-based negative regulation: Hfq and its accomplices. J. Biol. Chem. *288,* 7996-8003.

Desnoyers, G., and Masse, E. (2012). Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. Genes Dev. *26,* 726-739.

Durand, D., Vives, C., Cannella, D., Perez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010). NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. J. Struct. Biol. *169,* 45-53.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.P., and Ehresmann, B. (1987). Probing the structure of RNAs in solution. Nucleic Acids Res. *15,* 9109-9128.

Fang, X., Wang, J., O'Carroll, I.P., Mitchell, M., Zuo, X., Wang, Y., Yu, P., Liu, Y., Rausch, J.W., Dyba, M.A.*, et al.* (2013). An unusual topological structure of the HIV-1 Rev response element. Cell *155,* 594-605.

Figueroa-Bossi, N., Valentini, M., Malleret, L., Fiorini, F., and Bossi, L. (2009). Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target. Genes Dev. *23,* 2004-2015.

Folichon, M., Allemand, F., Regnier, P., and Hajnsdorf, E. (2005). Stimulation of poly(A) synthesis by Escherichia coli poly(A)polymerase I is correlated with Hfq binding to poly(A) tails. FEBS J. *272,* 454-463.

Foloppe, N., and MacKerell, J.,Alexander D. (2000). All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. Journal of Computational Chemistry *21,* 86-104.

Franze de Fernandez, M.T., Eoyang, L., and August, J.T. (1968). Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. Nature *219,* 588.

Franze de Fernandez, M.T., Hayward, W.S., and August, J.T. (1972). Bacterial proteins required for replication of phage Q ribonucleic acid. Pruification and properties of host factor I, a ribonucleic acid-binding protein. The Journal of Biological Chemistry *247,* 824.

Geissmann, T.A., and Touati, D. (2004). Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. EMBO J. *23,* 396-405.

Glatter, O., and Kratky, O. (1982). Small angle X-ray scattering (London: Academic).

Gottesman, S., McCullen, C.A., Guillier, M., Vanderpool, C.K., Majdalani, N., Benhammou, J., Thompson, K.M., FitzGerald, P.C., Sowa, N.A., and FitzGerald, D.J. (2006). Small RNA regulators and the bacterial response to stress. Cold Spring Harb. Symp. Quant. Biol. *71,* 1-11.

Gottesman, S., and Storz, G. (2011). Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb Perspect. Biol. *3,* 10.1101/cshperspect.a003798.

Guillier, M., and Gottesman, S. (2006). Remodelling of the Escherichia coli outer membrane by two small regulatory RNAs. Mol. Microbiol. *59,* 231-247.

Hammerle, H., Beich-Frandsen, M., Vecerek, B., Rajkowitsch, L., Carugo, O., Djinovic-Carugo, K., and Blasi, U. (2012). Structural and biochemical studies on ATP binding and hydrolysis by the Escherichia coli RNA chaperone Hfq. PLoS One *7,* e50892.

Hammerle, H., Vecerek, B., Resch, A., and Blasi, U. (2013). Duplex formation between the sRNA DsrA and rpoS mRNA is not sufficient for efficient RpoS synthesis at low temperature. RNA Biol. *10,* 1834-1841.

Henderson, C.A., Vincent, H.A., Casamento, A., Stone, C.M., Phillips, J.O., Cary, P.D., Sobott, F., Gowers, D.M., Taylor, J.E., and Callaghan, A.J. (2013). Hfq binding changes the structure of Escherichia coli small noncoding RNAs OxyS and RprA, which are involved in the riboregulation of rpoS. RNA *19,* 1089-1104.

Hengge-Aronis, R. (2002). Stationary phase gene regulation: what makes an Escherichia coli promoter sigmaS-selective? Curr. Opin. Microbiol. *5,* 591-595.

Hoe, C.H., Raabe, C.A., Rozhdestvensky, T.S., and Tang, T.H. (2013). Bacterial sRNAs: regulation in stress. Int. J. Med. Microbiol. *303,* 217-229.

Holmqvist, E., and Vogel, J. (2013). A small RNA serving both the Hfq and CsrA regulons. Genes Dev. *27,* 1073-1078.

Horstmann, N., Orans, J., Valentin-Hansen, P., Shelburne, S.A.,3rd, and Brennan, R.G. (2012). Structural mechanism of Staphylococcus aureus Hfq binding to an RNA A-tract. Nucleic Acids Res. *40,* 11023-11035.

Hura, G.L., Menon, A.L., Hammel, M., Rambo, R.P., Poole, F.L.,2nd, Tsutakawa, S.E., Jenney, F.E.,Jr, Classen, S., Frankel, K.A., Hopkins, R.C*., et al.* (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat. Methods *6,* 606-612.

Hussein, R., and Lim, H.N. (2011). Disruption of small RNA signaling caused by competition for Hfq. Proc. Natl. Acad. Sci. U. S. A. *108,* 1110-1115.

Ikeda, Y., Yagi, M., Morita, T., and Aiba, H. (2011). Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli. Mol. Microbiol. *79,* 419-432.

Ishikawa, H., Otaka, H., Maki, K., Morita, T., and Aiba, H. (2012). The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3' poly(U) tail. RNA *18,* 1062-1074.

Jorgensen, M.G., Thomason, M.K., Havelund, J., Valentin-Hansen, P., and Storz, G. (2013). Dual function of the McaS small RNA in controlling biofilm formation. Genes Dev. *27,* 1132-1145.

Jousselin, A., Metzinger, L., and Felden, B. (2009). On the facultative requirement of the bacterial RNA chaperone, Hfq. Trends Microbiol. *17,* 399-405.

Kadowaki, M.A., Iulek, J., Barbosa, J.A., Pedrosa Fde, O., de Souza, E.M., Chubatsu, L.S., Monteiro, R.A., de Oliveira, M.A., and Steffens, M.B. (2012). Structural characterization of the RNA chaperone Hfq from the nitrogen-fixing bacterium Herbaspirillum seropedicae SmR1. Biochim. Biophys. Acta *1824,* 359.

Kajitani, M., and Ishihama, A. (1991). Identification and sequence determination of the host factor gene for bacteriophage Q beta. Nucleic Acids Res. *19,* 1063-1066.

Kawamoto, H., Koide, Y., Morita, T., and Aiba, H. (2006). Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. Mol. Microbiol. *61,* 1013-1022.

Kim, N., Laing, C., Elmetwaly, S., Jung, S., Curuksu, J., and Schlick, T. (2014). Graph-based sampling for approximating global helical topologies of RNA. Proc. Natl. Acad. Sci. U. S. A. *111,* 4079-4084.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and and Svergun, D.I. (2003). *PRIMUS*: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography *36,* 1277-1282.

Kovach, A.R., Hoff, K.E., Canty, J.T., Orans, J., and Brennan, R.G. (2014). Recognition of U-rich RNA by Hfq from the Gram-positive pathogen Listeria monocytogenes. RNA

Laing, C., and Schlick, T. (2009). Analysis of four-way junctions in RNA structures. J. Mol. Biol. *390,* 547-559.

Laing, C., Wen, D., Wang, J.T., and Schlick, T. (2012). Predicting coaxial helical stacking in RNA junctions. Nucleic Acids Res. *40,* 487-498.

Le Derout, J., Boni, I.V., Regnier, P., and Hajnsdorf, E. (2010). Hfq affects mRNA levels independently of degradation. BMC Mol. Biol. *11,* 17-2199-11-17.

Lease, R.A., and Woodson, S.A. (2004). Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. J. Mol. Biol. *344,* 1211-1223.

Lescoute, A., and Westhof, E. (2006). Topology of three-way junctions in folded RNAs. RNA *12,* 83-93.

Leung, A.K.W., Nagai, K., and Li, J. (2011). Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. Nature *473,* 536.

Link, T.M., Valentin-Hansen, P., and Brennan, R.G. (2009). Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. Proc. Natl. Acad. Sci. U. S. A. *106,* 19292-19297.

Lipfert, J., Ouellet, J., Norman, D.G., Doniach, S., and Lilley, D.M. (2008). The complete VS ribozyme in solution studied by small-angle X-ray scattering. Structure *16,* 1357-1367.

Majdalani, N., Chen, S., Murrow, J., St John, K., and Gottesman, S. (2001). Regulation of RpoS by a novel small RNA: the characterization of RprA. Mol. Microbiol. *39,* 1382-1394.

Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T., and Gottesman, S. (1998). DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. Proc. Natl. Acad. Sci. U. S. A. *95,* 12462-12467.

Majdalani, N., Hernandez, D., and Gottesman, S. (2002). Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. Mol. Microbiol. *46,* 813-826.

Mandin, P., and Gottesman, S. (2010). Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. EMBO J. *29,* 3094-3107.

Mandin, P., and Gottesman, S. (2009). A genetic approach for finding small RNAs regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-component system. Mol. Microbiol. *72,* 551-565.

Masse, E., and Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *99,* 4620-4625.

McCullen, C.A., Benhammou, J.N., Majdalani, N., and Gottesman, S. (2010). Mechanism of positive regulation by DsrA and RprA small noncoding RNAs: pairing increases translation and protects rpoS mRNA from degradation. J. Bacteriol. *192,* 5559-5571.

McGinnis, J.L., Dunkle, J.A., Cate, J.H., and Weeks, K.M. (2012). The mechanisms of RNA SHAPE chemistry. J. Am. Chem. Soc. *134,* 6617-6624.

Michaux, C., Verneuil, N., Hartke, A., and Giard, J.C. (2014). Physiological roles of small RNA molecules. Microbiology *160,* 1007-1019.

Mika, F., Busse, S., Possling, A., Berkholz, J., Tschowri, N., Sommerfeldt, N., Pruteanu, M., and Hengge, R. (2012). Targeting of csgD by the small regulatory RNA RprA links stationary phase, biofilm formation and cell envelope stress in Escherichia coli. Mol. Microbiol. *84,* 51-65.

Mikulecky, P.J., Kaw, M.K., Brescia, C.C., Takach, J.C., Sledjeski, D.D., and Feig, A.L. (2004). Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. Nat. Struct. Mol. Biol. *11,* 1206-1214.

Moll, I., Afonyushkin, T., Vytvytska, O., Kaberdin, V.R., and Blasi, U. (2003). Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. RNA *9,* 1308-1314.

Moller, T., Franch, T., Hojrup, P., Keene, D.R., Bachinger, H.P., Brennan, R.G., and Valentin-Hansen, P. (2002). Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Mol. Cell *9,* 23-30.

Moon, K., and Gottesman, S. (2011). Competition among Hfq-binding small RNAs in Escherichia coli. Mol. Microbiol. *82,* 1545-1562.

Moore, S.D. (2011). Assembling new Escherichia coli strains by transduction using phage P1. Methods Mol. Biol. *765,* 155-169.

Morita, T., Maki, K., and Aiba, H. (2005). RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. Genes Dev. *19,* 2176-2186.

Moskaleva, O., Melnik, B., Gabdulkhakov, A., Garber, M., Nikonov, S., Stolboushkina, E., and Nikulin, A. (2010). The structures of mutant forms of Hfq from Pseudomonas aeruginosa reveal the importance of the conserved His57 for the protein hexamer organization. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. *66,* 760-764.

Muffler, A., Fischer, D., and Hengge-Aronis, R. (1996). The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for rpoS translation in Escherichia coli. Genes Dev. *10,* 1143.

Mura, C., Cascio, D., Sawaya, M.R., and Eisenberg, D.S. (2001). The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core. Proc. Natl. Acad. Sci. U. S. A. *98,* 5532.

Mura, C., Randolph, P.S., Patterson, J., and Cozen, A.E. (2013). Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. RNA Biol. *10,* 636-651.

Murina, V., Lekontseva, N., and Nikulin, A. (2013). Hfq binds ribonucleotides in three different RNA-binding sites. Acta Crystallogr. D Biol. Crystallogr. *69,* 1504-1513.

Murina, V.N., Melnik, B.S., Filimonov, V.V., Uhlein, M., Weiss, M.S., Muller, U., and Nikulin, A.D. (2014). Effect of conserved intersubunit amino acid substitutions on Hfq protein structure and stability. Biochemistry (Mosc) *79,* 469-477.

Murina, V.N., and Nikulin, A.D. (2011). RNA-binding Sm-like proteins of bacteria and archaea. similarity and difference in structure and function. Biochemistry (Mosc) *76,* 1434-1449.

Nielsen, J.S., Boggild, A., Andersen, C.B., Nielsen, G., Boysen, A., Brodersen, D.E., and Valentin-Hansen, P. (2007). An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from Methanococcus jannaschii. RNA *13,* 2213-2223.

Nikulin, A., Stolboushkina, E., Perederina, A., Vassilieva, I., Blaesi, U., Moll, I., Kachalova, G., Yokoyama, S., Vassylyev, D., Garber, M., and Nikonov, S. (2005). Structure of *Pseudomonas aeruginosa* Hfq protein. Acta Crystallographica: Section D (Wiley-Blackwell) *61,* 141.

Overgaard, M., Johansen, J., Moller-Jensen, J., and Valentin-Hansen, P. (2009). Switching off small RNA regulation with trap-mRNA. Mol. Microbiol. *73,* 790-800.

Panja, S., Schu, D.J., and Woodson, S.A. (2013). Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. Nucleic Acids Res.

Panja, S., and Woodson, S.A. (2012). Hfq proximity and orientation controls RNA annealing. Nucleic Acids Res. *40,* 8690-8697.

Parisien, M., and Major, F. (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature *452,* 51-55.

Peng, Y., Soper, T.J., and Woodson, S.A. (2014). Positional Effects of AAN Motifs in rpoS Regulation by sRNAs and Hfq. J. Mol. Biol. *426,* 275-285.

Peng, Y., Soper, T.J., and Woodson, S.A. (2012). RNase footprinting of protein binding sites on an mRNA target of small RNAs. Methods Mol. Biol. *905,* 213-224.

Perez-Salas, U.A., Rangan, P., Krueger, S., Briber, R.M., Thirumalai, D., and Woodson, S.A. (2004). Compaction of a bacterial group I ribozyme coincides with the assembly of core helices. Biochemistry *43,* 1746-1753.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., and and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. Journal of Applied Crystallography *45,* 342-350.

Petoukhov, M.V., and Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys. J. *89,* 1237-1250.

Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K.W., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. Nature *458,* 475.

Prevost, K., Salvail, H., Desnoyers, G., Jacques, J.F., Phaneuf, E., and Masse, E. (2007). The small RNA RyhB activates the translation of shiA mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. Mol. Microbiol. *64,* 1260-1273.

Rabhi, M., Espeli, O., Schwartz, A., Cayrol, B., Rahmouni, A.R., Arluison, V., and Boudvillain, M. (2011). The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. EMBO J. *30,* 2805-2816.

Rajkowitsch, L., and Schroeder, R. (2007). Coupling RNA annealing and strand displacement: a FRET-based microplate reader assay for RNA chaperone activity. BioTechniques *43,* 304, 306, 308 passim.

Rambo, R.P., and Tainer, J.A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. Nature *496,* 477-481.

Rambo, R.P., and Tainer, J.A. (2013). Super-Resolution in Solution X-Ray Scattering and Its Applications to Structural Systems Biology. Annual Review of Biophysics *42,* 415.

Ramos, C.G., Grilo, A.M., da Costa, P.J., Feliciano, J.R., and Leitao, J.H. (2013). MtvR is a global small noncoding regulatory RNA in Burkholderia cenocepacia. J. Bacteriol. *195,* 3514-3523.

Receveur-Brechot, V., and Durand, D. (2012). How random are intrinsically disordered proteins? A small angle scattering perspective. Curr. Protein Pept. Sci. *13,* 55-75.

Robinson, K.E., Orans, J., Kovach, A.R., Link, T.M., and Brennan, R.G. (2014). Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. Nucleic Acids Res. *42,* 2736-2749.

Salim, N.N., and Feig, A.L. (2010). An upstream Hfq binding site in the fhlA mRNA leader region facilitates the OxyS-fhlA interaction. PLoS One *5,* 10.1371/journal.pone.0013028.

Salvail, H., Caron, M.P., Belanger, J., and Masse, E. (2013). Antagonistic functions between the RNA chaperone Hfq and an sRNA regulate sensitivity to the antibiotic colicin. EMBO J. *32,* 2764-2778.

Sambrook, J., and Russell, D.W. (2006). Inverse PCR. CSH Protoc. *2006,* 10.1101/pdb.prot3487.

Sauer, E. (2013). Structure and RNA-binding properties of the bacterial LSm protein Hfq. RNA Biol. *10,* 610-618.

Sauer, E., Schmidt, S., and Weichenrieder, O. (2012). Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. Proc. Natl. Acad. Sci. U. S. A. *109,* 9396-9401.

Sauer, E., and Weichenrieder, O. (2011). Structural basis for RNA 3'-end recognition by Hfq. Proc. Natl. Acad. Sci. U. S. A. *108,* 13065-13070.

Sauter, C., Basquin, J., and Suck, D. (2003). Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. Nucleic Acids Res. *31,* 4091-4098.

Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2013). Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys. J. *105,* 962-974.

Schumacher, M.A., Pearson, R.F., Moller, T., Valentin-Hansen, P., and Brennan, R.G. (2002). Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. EMBO J. *21,* 3546-3556.

Schuppli, D., Miranda, G., Tsui, H.C., Winkler, M.E., Sogo, J.M., and Weber, H. (1997). Altered 3'-terminal RNA structure in phage Qbeta adapted to host factor-less Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *94,* 10239.

Senear, A.W., and Steitz, J.A. (1976). Site-specific interaction of Qbeta host factor and ribosomal protein S1 with Qbeta and R17 bacteriophage RNAs. The Journal of Biological Chemistry *251,* 1902.

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., Hinton, J.C., and Vogel, J. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet. *4,* e1000163.

Sledjeski, D.D., Whitman, C., and Zhang, A. (2001). Hfq is necessary for regulation by the untranslated RNA DsrA. J. Bacteriol. *183,* 1997-2005.

Sobrero, P., and Valverde, C. (2012). The bacterial protein Hfq: much more than a mere RNA-binding factor. Crit. Rev. Microbiol. *38,* 276-299.

Someya, T., Baba, S., Fujimoto, M., Kawai, G., Kumasaka, T., and Nakamura, K. (2012). Crystal structure of Hfq from Bacillus subtilis in complex with SELEX-derived RNA aptamer: insight into RNA-binding properties of bacterial Hfq. Nucleic Acids Res. *40,* 1856-1867.

Soper, T., Mandin, P., Majdalani, N., Gottesman, S., and Woodson, S.A. (2010). Positive regulation by small RNAs and the role of Hfq. Proc. Natl. Acad. Sci. U. S. A. *107,* 9602-9607.

Soper, T.J., Doxzen, K., and Woodson, S.A. (2011). Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. RNA *17,* 1544-1550.

Soper, T.J., and Woodson, S.A. (2008). The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. RNA *14,* 1907-1917.

Sorger-Domenigg, T., Sonnleitner, E., Kaberdin, V.R., and Blasi, U. (2007). Distinct and overlapping binding sites of Pseudomonas aeruginosa Hfq and RsmA proteins on the non-coding RNA RsmY. Biochem. Biophys. Res. Commun. *352,* 769-773.

Steen, K.A., Rice, G.M., and Weeks, K.M. (2012). Fingerprinting noncanonical and tertiary RNA structures by differential SHAPE reactivity. J. Am. Chem. Soc. *134,* 13160-13163.

Storz, G., Opdyke, J.A., and Zhang, A. (2004). Controlling mRNA stability and translation with small, noncoding RNAs. Curr. Opin. Microbiol. *7,* 140-144.

Storz, G., Vogel, J., and Wassarman, K.M. (2011). Regulation by small RNAs in bacteria: expanding frontiers. Mol. Cell *43,* 880-891.

Sukhodolets, M.V., and Garges, S. (2003). Interaction of Escherichia coli RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. Biochemistry *42,* 8022-8034.

Sun, X., Zhulin, I., and Wartell, R.M. (2002). Predicted structure and phyletic distribution of the RNA-binding protein Hfq. Nucleic Acids Res. *30,* 3662-3671.

Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. *76,* 2879-2886.

Svergun, D.I., Barberato C., and and Koch M. H. J. (1995). CRYSOL - a Program to Evaluate X-ray Solution Scattering of biological Macromolecules from Atomic Coordinates. Journal of Applied Crystallography *28,* 768-773.

Tan, E.M., and Kunkel, H.G. (1966). Characteristics of a soluble nuclear antigen precipitating with sera of patients with systemic lupus erythematosus. J. Immunol. *96,* 464-471.

Teitel, J.M. (2000). Clinical approach to the patient with unexpected bleeding. Clin. Lab. Haematol. *22 Suppl 1,* 9-11; discussion 30-2.

Tharun, S., He, W., Mayes, A.E., Lennertz, P., Beggs, J.D., and Parker, R. (2000). Yeast Sm-like proteins function in mRNA decapping and decay. Nature *404,* 515-518.

Thomason, M.K., Fontaine, F., De Lay, N., and Storz, G. (2012). A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in Escherichia coli. Mol. Microbiol. *84,* 17-35.

Thore, S., Mayer, C., Sauter, C., Weeks, S., and Suck, D. (2003). Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. J. Biol. Chem. *278,* 1239-1247.

Timmermans, J., and Van Melderen, L. (2010). Post-transcriptional global regulation by CsrA in bacteria. Cellular and Molecular Life Sciences: CMLS *67,* 2897.

Tompa, P., and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB J. *18,* 1169-1175.

Törö, I., Thore, S., Mayer, C., Basquin, J., Séraphin, B., and Suck, D. (2001). RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. EMBO J. *20,* 2293.

Tsui, H.C., Leung, H.C., and Winkler, M.E. (1994). Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. Mol. Microbiol. *13,* 35.

Updegrove, T., Wilf, N., Sun, X., and Wartell, R.M. (2008). Effect of Hfq on RprA-rpoS mRNA pairing: Hfq-RNA binding and the influence of the 5' rpoS mRNA leader region. Biochemistry *47,* 11184-11195.

Updegrove, T.B., and Wartell, R.M. (2011). The influence of Escherichia coli Hfq mutations on RNA binding and sRNA*mRNA duplex formation in rpoS riboregulation. Biochim. Biophys. Acta *1809,* 532-540.

Urban, J.H., and Vogel, J. (2007). Translational control and target recognition by Escherichia coli small RNAs in vivo. Nucleic Acids Res. *35,* 1018-1037.

V. V. Volkov and D. I. Svergun. (2003). **Uniqueness of *ab initio* shape determination in small-angle scattering**. Journal of Applied Crystallography *36,* 860-864.

Vecerek, B., Moll, I., Afonyushkin, T., Kaberdin, V., and Blasi, U. (2003). Interaction of the RNA chaperone Hfq with mRNAs: direct and indirect roles of Hfq in iron metabolism of Escherichia coli. Mol. Microbiol. *50,* 897-909.

Vecerek, B., Rajkowitsch, L., Sonnleitner, E., Schroeder, R., and Blasi, U. (2008). The C-terminal domain of Escherichia coli Hfq is required for regulation. Nucleic Acids Res. *36,* 133-143.

Viegas, S.C., Pfeiffer, V., Sittka, A., Silva, I.J., Vogel, J., and Arraiano, C.M. (2007). Characterization of the role of ribonucleases in Salmonella small RNA decay. Nucleic Acids Res. *35,* 7651-7664.

Vogel, J., and Luisi, B.F. (2011). Hfq and its constellation of RNA. Nat. Rev. Microbiol. *9,* 578-589.

Vytvytska, O., Moll, I., Kaberdin, V.R., von Gabain, A., and Blasi, U. (2000). Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding. Genes Dev. *14,* 1109-1118.

Wang, W., Wang, L., Wu, J., Gong, Q., and Shi, Y. (2013). Hfq-bridged ternary complex is important for translation activation of rpoS by DsrA. Nucleic Acids Res. *41,* 5938-5948.

Wang, W., Wang, L., Zou, Y., Zhang, J., Gong, Q., Wu, J., and Shi, Y. (2011). Cooperation of Escherichia coli Hfq hexamers in DsrA binding. Genes Dev. *25,* 2106-2117.

Waters, L.S., and Storz, G. (2009). Regulatory RNAs in bacteria. Cell *136,* 615-628.

Watson, M.C., and Curtis, J.E. (2014). Probing the average local structure of biomolecules using small-angle scattering and scaling laws. Biophys. J. *106,* 2474-2482.

Weichenrieder, O. (2014). RNA binding by Hfq and ring-forming (L)Sm proteins: a trade-off between optimal sequence readout and RNA backbone conformation. RNA Biol. *11,* 537-549.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006). Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat. Protoc. *1,* 1610-1616.

Will, C.L., and Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. Curr. Opin. Cell Biol. *13,* 290-301.

Yakhnin, A.V., Baker, C.S., Vakulskas, C.A., Yakhnin, H., Berezin, I., Romeo, T., and Babitzke, P. (2013). CsrA activates flhDC expression by protecting flhDC mRNA from RNase E-mediated cleavage. Mol. Microbiol. *87,* 851-866.

Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C.C., Pettersen, E.F., Goddard, T.D., Meng, E.C., Sali, A., and Ferrin, T.E. (2012). UCSF Chimera, MODELLER, and IMP: an integrated modeling system. J. Struct. Biol. *179,* 269-278.

Yonekura, K., Watanabe, M., Kageyama, Y., Hirata, K., Yamamoto, M., and Maki-Yonekura, S. (2013). Post-transcriptional regulator Hfq binds catalase HPII: crystal structure of the complex. PLoS One *8,* e78216.

Zaug, A.J., Grosshans, C.A., and Cech, T.R. (1988). Sequence-specific endoribonuclease activity of the Tetrahymena ribozyme: enhanced cleavage of certain oligonucleotide substrates that form mismatched ribozyme-substrate complexes. Biochemistry *27,* 8924-8931.

Zhang, A., Altuvia, S., Tiwari, A., Argaman, L., Hengge-Aronis, R., and Storz, G. (1998). The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein. EMBO J. *17,* 6061-6068.

Zhang, A., Schu, D.J., Tjaden, B.C., Storz, G., and Gottesman, S. (2013). Mutations in interaction surfaces differentially impact E. coli Hfq association with small RNAs and their mRNA targets. J. Mol. Biol.

Zhang, A., Wassarman, K.M., Ortega, J., Steven, A.C., and Storz, G. (2002). The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Mol. Cell *9,* 11-22.

Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G., and Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. Mol. Microbiol. *50,* 1111.

Zhang, Y., and Hong, G. (2009). Post-transcriptional regulation of NifA expression by Hfq and RNase E complex in Rhizobium leguminosarum bv. viciae. Acta Biochim. Biophys. Sin. (Shanghai) *41,* 719-730.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31,* 3406-3415.

Curriculum Vitae

Yi (Penny) Peng

3811 Canterbury Rd., Apt. 605, Baltimore, MD 21218 | Mobile: 443-691-2358 |
pengyi.0801@gmail.com

Education

Ph.D. Candidate in Biology, **Johns Hopkins University**, U.S., 2010-2014

B.S. in Biochemistry (Minor in Biology), **Hong Kong University of Science and Technology**, H.K., 2007-2010

- First Class Honors degree.
- GPA: 11.18/12; GRE: Q 800, V 600, A 3.0

Exchange program, **University of Waterloo**, Canada, Sep-Dec 2008

- GPA: 3.9/4.0

B.S. in Biology (transferred before finishing degree), **Fudan University**, China, 2006-2007

- Grade: A (top 10%)

Honors and Awards

- International Biology Olympiad Competition Gold Medal (third-place awardee among 185 participants from 48 countries), Río Cuarto, Argentina, Jul 2006.
- Outstanding Freshman Scholarship First Class (RMB ¥10,000), Fudan U., 2006
- Dean’s List, HKUST, 2007-2010
- School of Science Scholarship (HKD $40,000/yr), HKUST, 2007-2010
- D H Chen Foundation Scholarship (HKD $40,000/yr), HKUST, 2007-2010
- Academic Achievement Medal (top 5%), HKUST, 2010
- Summer research fellowship (25 fellowship awardees selected from over 500 applications around the globe, CHF 4200), EPFL, 2009

Experience

Freelance translator, CACTUS, Nov 2014-present

- Translating academic and scientific manuscripts, conference proceedings, and presentations across a range of subjects
- Specializing in Chinese-English translation for biomedical and molecular biology subjects

Health Information Technology Consultant Intern, Technical Consulting & Research, Inc., Oct 2014-present

- Assisting clients to understand and implement Electronic Health Records software

Research associate, Johns Hopkins University, May 2011-Sep 2014

- Thesis research on "structural models of bacterial Hfq protein in complex with RNAs" with Dr. Sarah A. Woodson in biophysics department.
- Applied genetic assays to measure mRNA translation levels *in vivo.*
- Utilized molecular and biophysical approaches to characterize RNA-protein interactions *in vitro.*
- Developed biochemical and computational methods to probe mRNA structural change upon protein binding.

Teaching assistant, Johns Hopkins University, Sep 2011-May 2012

- Led weekly 3-hour lab sessions for undergraduate biochemistry and molecular biology courses.
- 100% pass rate for the lab exams and average 10% higher score than parallel sessions.

Research associate, Hong Kong University of Science and Technology, Sep 2009-May 2010

- Thesis research on "regulations of synaptic complex assemblies" with Dr. Mingjie Zhang in biochemistry department.
- Developed yeast two-hybrid systems to study protein-protein interactions.

Summer research fellow, École polytechnique fédérale de Lausanne, Jul-Aug 2009

- Study “transcription factors in adipogenesis” with Dr. Bart Deplancke in bioengineering department.
- Established *in vivo* screening methods to identify novel adipogenesis regulators.

Summer research intern, Hong Kong University of Science and Technology, Jun-Aug 2008

- Study “regulatory mechanisms of neuronal differentiation” with Dr. Nancy IP in biochemistry department.
- Used co-immunoprecipitation methods to identify binding targets of signaling regulator CDK5 during dendrite development.

Part-time Biology Olympiad Competition coach, Beijing Qingbei Xuetang Education Technology Co., Mar-May 2007

- Created and supervised 2-month study plan for ~100 students at different levels to prepare for National Biology Olympiad Competition.
- Awarded with bonus compensation based on colleague evaluation and student feedback.

Publications and Presentations

- Peng Y.; Curtis,J.E., Fang,X.Y., and Woodson,S.A. “Structural model of an mRNA in complex with the bacterial chaperone Hfq”. *Proceedings of the National Academy of Sciences*. 2014; [Epub ahead of print].
- Peng,Y.; Soper,T.J., and Woodson,S.A. “Positional Effects of AAN Motifs in rpoS Regulation by sRNAs and Hfq”. *Journal of Molecular Biology*. 2014. 426: 275-285. (Featured cover story)
- Peng,Y.; Soper,T.J., and Woodson,S.A. “RNase footprinting of protein binding sites on an mRNA target of small RNAs”. *Methods Molecular Biolology.* 2012. 905: 213-224.

- Yi Peng. “Structural models of Hfq in complex with *rpoS* mRNA”. RNA society meeting oral presentation. Quebec, Canada. 2014.
- Yi Peng. “Structural models of Hfq in complex with *rpoS* mRNA”. NIH RNA club seminar talk. Bethesda, MD, U.S.A. 2014.
- Yi Peng. “Structural models of Hfq in complex with *rpoS* mRNA”. NIH RNA club seminar talk. Bethesda, MD, U.S.A. 2014.
- Yi Peng and Sarah A. Woodson. “Positions of the A-rich and U-rich sequences on the *rpoS* leader determine Hfq’s chaperone function”. Poster presentation. *International Conference on “Regulating with RNAs in Bacteria”*. Würzburg, Germany. 2013.
- Yi Peng and Sarah A. Woodson. “Positions of the A-rich and U-rich sequences on the rpoS leader determine Hfq’s chaperone function”. Poster presentation (awarded with “Best Student Poster Presentation”). *CMDB retreat*. Baltimore MD, U.S.A. 2012.

Interests and Activities

- Love swimming and jogging
- Have been and will continue traveling to many countries
- Enjoy watching documentaries on technology and social science

Language

English (fluent), Mandarin (native), and Cantonese (fluent)